Studies of the hindgut and faecal volatile organic compound metabolome and microbiome of the horse

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Rachael Siân Slater

November 2019

‘Most large animals, in fact are not the single individuals they seem to be. They are walking menageries, whole communities of different species which, in their various ways, are committed by evolution, for better or for worse, in sickness and in health, to live together'.

D. Attenborough, (1990)

Contents

List of Figures

List of Tables

List of Abbreviations

AF – Anaerobic fungi

ANOVA – Analysis of variance

AT – All tapeworm

AUC – Area under curve

BER – Balanced error rate

bp – Base pairs

CAR-PDMS – Carboxen-polydimethylsiloxane

CLR – Centered log ratio

CO – Control

CoV – Coefficient of variation

DIABLO - Data Integration Analysis for Biomarker discovery

DNA – Deoxyribonucleic acid

dNTP - Deoxyribonucleotide triphosphate

DVB-CAR-PDMS – Divinylbenzene-carboxen-polydimethylsiloxane

EGS – Equine grass sickness

ELISA - Enzyme-linked immunosorbent assay

ER – Error rate

FD – Freeze-dried

FDR – False discovery rate

FEC – Faecal egg count

FF – Facultative anaerobic fungi

GC – Gas chromatograph

GCMS – Gas chromatography mass spectroscopy

HS – Headspace

HSC – high strongyle count

HS-SPME-SPME – headspace solid-phase microextraction

ITS - Internal transcribed spacer

LASSO - Least absolute shrinkage and selection operator

LCV – Large colon volvulus

LME – Linear mixed effects modelling

LSC – Low strongyle count

LVC – Left ventral colon

MH – Medium and high tapeworm

MRT – Mean retention time

MS – Mass spectrometer

NFD – Non-freeze-dried

NMDS - Non-metric multidimensional scaling

NMR - Nuclear magnetic resonance

OTU – Operational taxonomic unit

PCA – Principal component analysis

PCR – Polymerase chain reaction

PERMANOVA – Permutational multivariate analysis of variance

PLS-DA – partial least squares discriminant analysis

Qiime – Quantitative Insights Into Microbial Ecology

qPCR – Quantitative polymerase chain reaction

ROC – Receiver operator characteristic

18S/16S rRNA – Svedberg ribosomal nucleic acid

SCFA – Short chain fatty acid

SCOD - Simple colonic obstruction and distension

sPLS-DA – Sparse partial least squares discriminant analysis

SPME – Solid-phase microextraction

TSS – Total sum scaling

TRFLP - Terminal restriction fragment length polymorphism

VFA – Volatile fatty acid

VOC - volatile organic compound

Abstract

The horse is a hindgut fermenter which relies on microbial digestion to provide more than half of its energy requirements. Disturbances to the microbiota can lead to colic (abdominal pain), diarrhoea and other disorders in horses. Advances in techniques to characterise the equine gut microbiome have revealed that this is a complex population and many factors are thought to contribute towards the composition of species present. Epidemiological studies have identified various horse and management risk factors for colic including season (and associated management changes), foaling and tapeworm infections. The contribution of the intestinal microbiota to the development of colic in relation to these risk factors is unknown. The aim of this thesis is to investigate the association of these factors with the intestinal (or faeces as a proxy) microbiome and metabolome. Few studies to date have attempted to understand the equine gut mycobiome and functional equine microbiome (metabolome). The faecal metabolome may provide simple, cost effective markers for microbial shifts that may be associated with equine disease including increased likelihood of colic.

A method to extract volatile organic compounds (VOCs) from equine faeces for gas chromatography mass spectrometry (GCMS) was developed and applied to longitudinal studies in healthy horses. The VOC profile and mycobiome (18S rRNA and internal transcribed spacer region 1 DNA sequences) of horses grazing at pasture over 12 months was characterised as well as the VOC profile of mares around the time of foaling. The optimised extraction method (with some modifications) and 16S rRNA gene sequencing was also applied to hindgut and rectal contents collected from horses identified as either positive or negative for *Anoplocephala perfoliata* (tapeworm) at post-mortem.

Horses grazing at pasture over 12 months showed a marked difference in VOC profile over time. Notably, there was a marked increase in fungal compounds and a shift in fungal species was observed in the autumn, a high-risk time of year for both colic and tapeworm infection. The VOC profile of periparturient mares remained stable throughout the sampling period, providing baseline data to assist identification of mares at increased likelihood of developing colic. Variations in the intestinal VOC profile and bacterial microbiome between horses with and without *A. perfoliata* infection were observed, even without the ability to control for extrinsic factors such as diet and other management factors. Furthermore, a combination of 10 VOCs emitted from rectal contents generated a sensitivity of 80% and a specificity of 70% when the model was tested on a sub-set of

samples. These pilot results warrant further study of parasites and their interactions with the equine intestinal microbiome using controlled study designs. Where multiple -omics were used, correlations of VOCs with both bacterial microbiome and mycobiome data were observed. These data indicate that VOCs have the potential to act as cost-effective markers for microbial shifts in horse intestinal contents and faeces.

Publications arising from this thesis

Poster

Federation of European Microbiological Societies (FEMS), Glasgow, UK, A comparison of the intestinal microbiome and metabolome of *Anoplocephala perfoliata* infected and non-infected horses: a pilot study. Hough R, Frau A, Hodgkinson J, Kenny J, Archer D, Probert C, 2019

Oral presentations

Royal Society of Chemistry GASG meeting for early career researchers, *Milton Keynes*, *UK,* "Equine faecal VOCs: methods to extract and biomarkers for tapeworm infection". Hough R, Archer D, Probert C, March 2018

British Society of Animal Science (BSAS), Chester, UK, "Temporal changes in the faecal volatile organic compound (VOC) metabolome of healthy mares pre- and post-partum". Hough R, Salem S, Archer D, Probert C, April 2016

Journal articles

Salem S.E., Hough, R., Probert C., Maddox T.W., Berg A., Antczak P., Ketley J.M. , Williams N.J., Stoneham SJ, Archer DC. (2019) A longitudinal study of the faecal microbiome and metabolome of periparturient mares, *PeerJ*.

Hough, R., Archer, D. & Probert, C. (2018) A comparison of sample preparation methods for extracting volatile organic compounds (VOCs) from equine faeces using HS-SPME. *Metabolomics*, 14, 19.

Acknowledgements

First of all, I would like to thank my primary supervisor Prof. Chris Probert for taking on (and remaining enthusiastic about!) a horse project and for finding ways to support me as a self-funded student. I feel very lucky that my application reached your desk. Another big thank you goes to my secondary supervisor Prof. Debbie Archer, for your enthusiasm, support and those long trips down to the abattoir! I really couldn't have asked for a better supervisory team.

My sincerest thanks go to Dr. Shebl Salem for the longitudinal study samples (which you worked tirelessly to collect) and for your help and advice with analysis. Thank you to Dr. Alessandra Frau for your amazing support and advice in the lab, with bioinformatics and for being a great friend. It really has been fantastic to have worked alongside two great role models during my PhD.

When I first joined the lab, I received a warm welcome from Sophie, Arno, Rapha and Tan. For this, and for introducing me to the world of metabolomics and R I am so grateful, thank you. I would also like to thank Dr. David Hughes for your statistical advice, particularly for managing longitudinal multivariate datasets. A huge thank you is also owed to Dr. Elinor Chapman for your support and kindness, particularly in the weeks leading up to my viva.

Thank you to my parents for your unfaltering support and for always letting me find my own path. Finally, I would like to thank my husband, James for being by my side every step of the way and sharing the ups and downs with me.

Diolch yn fawr iawn pawb.

Chapter 1 Introduction

The intestinal microbiota plays an important role in health and disease of the horse. This chapter provides an overview of current knowledge of the equine microbiota in terms of how it contributes towards the health of the host, the factors that affect the composition of species present and how it is associated with gastrointestinal disease. Work presented in this thesis was designed to address some identified gaps in knowledge regarding the equine microbiome in health and disease, including horses with tapeworm infections. This work involved investigation of analysis of volatile organic compounds (VOCs) (the functional microbiome) as a proxy for hindgut microorganisms. Improved knowledge about the equine intestinal microbiota in healthy horses, changes in disease and use of VOCs may lead to better strategies to prevention of gastrointestinal diseases including colic and potential biomarkers and therapies for parasites.

1.1 Anatomy and physiology of the equine gastrointestinal tract

The horse (this includes other equids such as ponies and donkeys) is a monogastric, herbivore that consumes a largely fibrous diet and would naturally graze for around 18 hours per day (Rubenstein, 1981). The gastrointestinal tract of the horse has numerous adaptations to meet nutritional requirements. The hindgut of the horse provides fermentation chamber-like compartments where an active community of microbes perform the breakdown of fibrous fractions of the diet into absorbable substrates. For this reason, the horse is referred to as a hindgut fermenter (Argenzio & Hintz, 1970). A complete schematic diagram of the equine gastrointestinal tract is shown in Figure 1.1.

1.1.1 The Mouth

The mouth is where selection and intake of food occurs, and digestion is initiated. The horse uses its lips to pick up and separate food. Cutting of food is performed by the front incisor teeth before being relayed by the tongue to the molars and premolars for grinding. A bolus is then formed along the surface of the tongue which passes to the soft palate for swallowing (Frandson, 2003).

Saliva is produced by the salivary glands with the primary functions to moisten food for bolus formation for ease of swallowing, initiate digestion and buffer acid on arrival of the bolus at the stomach (Frape, 2010). The salivary glands of the horse are the mandibular, parotid,

sublingual and buccal glands (Dyce *et al.*, 2018). Saliva is produced in response to numerous triggers including: mechanical action of mastication, taste and smell of food as well as pain, infection and exertion (Hector & Linden, 1999; Proctor & Carpenter, 2007; Ligtenberg *et al.*, 2015).

Potassium and bicarbonate are present as buffers in equine saliva although amylase is not present, which is an indicator the horse has not evolved to consume a high starch diet (Ellis & Hill, 2005). A unique protein, known as latherin, is found in horse sweat and saliva (McDonald *et al.*, 2009). The function of latherin is thought to increase the surface spread of saliva to allow it to coat fibrous material more readily for mastication and digestion (McDonald *et al*., 2009). It has also been suggested that latherin reduces microbial adherence to teeth and oral surfaces (Kennedy, 2011).

1.1.2 The Oesophagus and Stomach

The oesophagus is a muscular tube along which food passes from the mouth to the stomach by peristaltic contractions. The structure and function of the oesophagus is relatively uniform across mammalian species. The oesophagus is made up of four layers; an outer fibrous protective sheath, a layer of circular and striated muscle, a sub-mucosal layer and the innermost layer of mucosa lined with stratified epithelial cells (Ellis & Hill, 2005). A cardiac-sphincter valve is located where the oesophagus meets the stomach (Dyce *et al.*, 2018). The probable function of this valve is to prevent gastric reflux and eructation of gases, possibly to prevent the acidic contents of the stomach from escaping and causing damage to the upper part of the tract as chronic gastric reflux can lead to severe ulceration (Long & Orlando, 1999; Heidmann *et al.*, 2004; Ellis & Hill, 2005). A second valve is situated where the small intestine joins the stomach, known as the pyloric sphincter. This valve regulates the flow of digesta from the stomach into the small intestine. The stomach, a J shaped organ, contributes to 10% of the total volume of the tract. In comparison to the overall size of the horse this is relatively small. The stomach is split into four regions, which are glandular - pyloric and fundic and non-glandular - oesophageal and cardiac (Merrit, 1999).

1.1.3 Small intestine

The small intestine is around 21-25m in length in an average sized 450kg horse, which is 30% of the entire gastrointestinal tract. It consists of three regions; duodenum, jejunum and ileum. Pancreatic juice containing digestive enzymes is secreted continuously into the duodenum and increases the luminal pH to 7 which provides the ideal environment for

mammalian enzymatic digestion. Mammalian enzymatic digestion occurs in the small intestine and is similar to other monogastric animals (McDonald *et al.*, 2011). In the horse, complex carbohydrate digestion is limited in the small intestine because of low concentrations of starch digesting enzymes including α-amylase. Therefore, complex carbohydrates may pass undigested into the hindgut (Richards *et al.*, 2004). Fibrous components of plant cell walls cannot be broken down by mammalian enzymes and must pass through into the hindgut for microbial digestion.

1.1.4 The hindgut

The large intestine is made up of the caecum (a fermentation chamber) and the large and small colon before terminating at the rectum, the site of faecal ball formation. Digesta spends the majority of time within the hindgut, which is usually around 35 hours (75-85% of total mean retention time) (Van Weyenberg *et al.*, 2006). Mean retention time (MRT) positively correlates with the degree of fibre breakdown. The caecum has been found to have the shortest MTR, in comparison to the colon, with the left ventral colon (LVC) having the longest MRT (Miyaji *et al.*, 2008b). This is reflected in fibre breakdown being more extensive in the ventral colon than the caecum (Miyaji *et al.*, 2008a).

In the adult horse the caecum is approximately 1 metre in length and has a capacity of 30 litres (Dyce *et al.*, 2018). The caecum is a sacculated structure with four distinct teniae which acts as support and in peristaltic contractions. The caecum is a major site of water absorption, but water is retained for longer in the proximal colon. Water exchange via the colonic mucosa, controlled by changes in osmotic gradient, is likely to be associated with the water demands of microbial fermentation (Argenzio *et al.*, 1974a).

On exiting the caecum, digesta enters the right and left ventral colons, the largest of the fermentation chambers of the tract. The narrow pelvic flexure, which slows down passage rate, joins the ventral colons to the left and right dorsal colons.

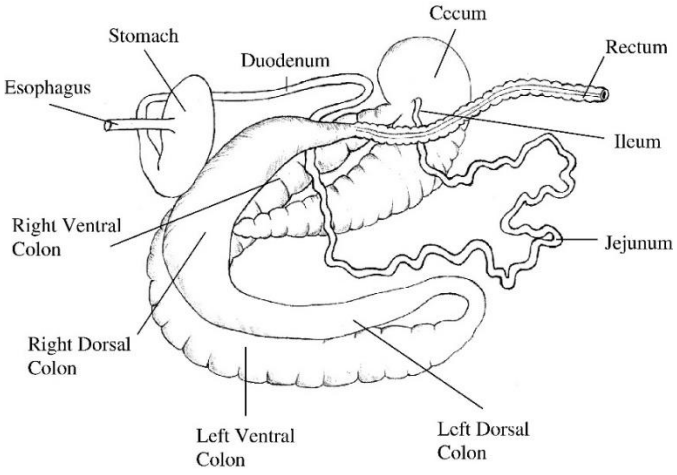


Figure 1.1 Digestive tract of the equine. Taken from (Van Weyenberg et al., 2006). Licence to re-use obtained from Elsevier (licence number: 4672530416670).

1.2 Microbiology of the gastrointestinal tract

A synergistic relationship exists between the gastrointestinal bacteria and host. The host provides an anaerobic environment with a neutral pH (hindgut regions) and a constant stream of substrate to 'feed' the bacteria. In return microbes digest complex polysaccharides that are indigestible by mammalian enzymes (Harris and Geor, 2009). The endogenous protein utilised by the microbes is converted into ammonia and microbial protein, largely for their own use, but the ammonia also aids water and sodium conservation for the host. The volatile fatty acids (VFAs) produced by microbes also provide energy for water and sodium absorption across host epithelial cells (Stevens and Hume, 1998). Furthermore, colonic blood flow is increased by luminal VFAs (Von Engelhardt *et al.*, 1989). Additional benefits of VFAs to the host include acting as a barrier against pathogens for the prevention of disorders during normal feeding conditions (Julliand, 1998).

1.2.1 Nutritional contribution

The hindgut is the major site of fibre breakdown by fermentation in the horse (Kern *et al.*, 1974). Fermentation is carried out by microbial communities present in these regions of the tract, producing VFAs of the short chain variety (fewer than six carbons). It is estimated that up to 1g of VFAs are produced per kg of horse bodyweight (Frape, 2010). The predominant VFAs formed are acetic acid, propanoic acid and butanoic acid (Daly *et al.*, 2001; Kristoffersen *et al.*, 2016). Additional products including lactic acid, formic acid, ethanol and succinic acid are also made. Furthermore, fermentation creates ammonia, hydrogen and carbon dioxide as by-products. Cellulolytic bacteria produce VFAs via extracellular cellulase and other

enzymes to degrade cellulose and hemicellulose into simple sugars. The Embden-Meyerhof-Parnas pathway is the main route of the conversion of glucose into pyruvate (Miller & Wolin, 1996). Pyruvate may then be utilised to produce various VFAs.

Proteolysis produces amino acids, phenols, ammonia and sulphides (Windey *et al.*, 2012). In the ruminant other fatty acids are formed from the deamination of amino acids by the microbiota producing: 2-methylpropanoic acid from valine, valeric acid from proline, butanoic acid, 2-methyl- from isoleucine and butanoic acid, 3-methyl- from leucine (McDonald *et al.*, 2011). While it is expected that similar occurs in the equine hindgut, this may be at a lesser extent than in the ruminant as the majority of proteolytic bacteria are present in the small intestine of the equid (Mackie & Wilkins, 1988).

Volatile fatty acids are absorbed across the intestinal mucosa and are a major source of energy for the horse (Bergman, 1990). The rate of absorption is influenced by chain length and therefore butanoic acid is absorbed first followed by propanoic acid and then acetic acid (Krehbiel, 2014). After absorption propanoic acid is converted into glucose in the liver and has been reported to account for 50% (hay diet) and 61% (hay and bran diet) of blood glucose in ponies (Simmons & Ford, 1991). Propanoic acid can also be stored as glycogen in the liver. Acetic and butanoic acids contribute towards fat synthesis and as energy sources for tissues (Von Engelhardt *et al.*, 1989). Butanoic acid is also used as an energy substrate by the intestinal epithelium. Bacterial acetic acid producers include: *Bifidobacteria*, *Eubacteria*, *Propionibacteria*, *Selenomonas* and Streptococci (Mackie & Wilkins, 1988). Butanoic acid producing organisms include *Roseburia* spp., *Eubacteria* and *Clostridia* (Leser & Mølbak, 2009; Louis & Flint, 2009). The proportions of VFAs produced in the hindgut are greatly affected by diet and the bacterial species present, as discussed in section 1.4.3.

1.2.2 Competitive pathogenicity

Commensal bacteria of the gastrointestinal tract have a symbiotic relationship with the host. The host provides an environment that is at a near neutral pH, anaerobic, thermostable and provides protein from sources including: alimentary protein, recycled urea and endogenous protein from cell desquamation for growth (Stevens & Hume, 1998; Santos *et al.*, 2010). Commensal gut bacteria also create a barrier to protect against the invasion of pathogenic bacteria. Numerous studies have concluded that a species-rich environment is beneficial for resisting pathogen invasion (Law & Morton, 1996; Shurin, 2000; Dillon *et al.*, 2005). It is thought that the microorganisms carry out this function by stimulating the immune system as well as competing for nutrition and epithelial binding sites (Naidu *et al.*, 1999). Examples

of these activities by bacteria include: production of short chain fatty acids by mixed cultures of *Veillonella* spp. (present in the equine stomach) and *Bacteroides fragilis* which inhibit the growth of pathogenic bacteria including *Salmonella enteritidis*, *Salmonella enterica* Typhimurium, *Escherichia coli* and *Pseudomonas aeruginosa* (Hinton & Hume, 1995). *B. fragilis* ferments glucose to form succinic acid which is then decarboxylated by *Veillonella* spp to produce propanoic acid, which lowers the pH and controls the growth of pathogenic species. The bacterium *Lactobacillus reuteri* Lr1, isolated from the equine gastrointestinal tract was found to reduce the adherence of *Clostridioides (Clostridium) difficile* C6 a bacterium associated with colitis, from 60% to 3% (Dicks *et al.*, 2015). Further discussion of the interaction of probiotics and the equine hindgut microbiota is in section 1.4.10.

1.2.3 Stimulation of the immune system

The gut microbiome plays a key role in mammalian health, which has emerged since the coding of the mammalian genome. The mammalian genome alone is not equipped to carry out all functions to maintain health and therefore requires the assistance of other organisms (Round & Mazmanian, 2009). In germ-free rodent studies, the introduction of commensal bacteria to the gastrointestinal tract has been found stimulate the immune system (Cebra, 1999). Germ-free animals have also been found to have slower gut epithelial turnover and decreased numbers of goblet and endocrine cells, indicating that microbes play a role in gut mucosal function (Savage *et al.*, 1981; Bates *et al.*, 2006). Microbial-produced VFAs are important components for intestinal barrier health and for inducing systemic immune responses. Their mechanism of signalling is via G-protein-coupled-receptors and they induce responses including: increased mucus secretion, protection against inflammation by regulating regulatory T-cells (Tregs) and promoting autophagy (Spiljar *et al.*, 2017). In mice fed an acetic acid supplemented diet, decreased lipid accumulation resulted in improved hepatic function and reduced body adiposity (Sahuri-Arisoylu *et al.*, 2016). Furthermore, a fibre-deprived diet resulted in colonic erosion of the mucus barrier in a gnotobiotic (colonised with human gut microbiota) mouse model (Desai *et al.*, 2016). Therefore, a stable gut microflora (maintained by a high fibre diet) may have important consequences for the overall health of the animal. In the horse, studies to investigate the stimulation of the immune system by the intestinal microbiota are yet to be performed. The immune response in relation to helminth infection and the intestinal microbiota has been investigated in a cohort of ponies and is discussed further in section 1.5.2 (Clark *et al.*, 2018). However, the specific mechanisms that underlie microbiota and immune system interactions are currently unknown in the horse.

1.3 The microorganisms of the equine gastrointestinal tract

The types of microbes present in the equine gastrointestinal tract include bacteria, fungi, protozoa and archaea (Orpin, 1981). The majority of bacteria are present in the hindgut (Table 1.1) and production of VFAs are greatest in the hindgut (Fombelle *et al.*, 2003). Some fermentation takes place in other regions of the tract including the stomach. The non-glandular regions of the stomach are populated with Lactobacilli bacteria which adhere to the stratified squamous epithelium cell surfaces (Yuki *et al.*, 2000). Further work has found other species of bacteria to be present in the stomach including: Streptococci, *Veillonella* and Actinobacilli (St-Pierre *et al.*, 2013). However it was concluded that since cellulolytic species of bacteria were not detected in the stomach, the function of the stomach is to regulate the rate of passage rather than a primary role of plant digestion and that the species of bacteria present provide a competitive environment for pathogenic species (St-Pierre *et al.*, 2013). In comparison to the large intestine, the stomach has fewer bacterial numbers (Table 1.1). Although some VFA absorption occurs in the stomach (pyloric region), it is not to the extent of that in the large intestine (Argenzio *et al.*, 1974b). Bacteria found in the lumen and mucosa of the small intestine are largely proteolytic and total bacterial numbers are lower than in the large intestine (Mackie & Wilkins, 1988). Bacterial species isolated in low numbers from the small intestine include: *Clostridium*, *Pseudomonas*, *Candida*, *Proteus* and *Staphylococcus* spp. (Julliand, 2005). The core bacterial community of the small intestine is also larger (32%) than the large intestine (5-15%) indicating a less diverse community in the small intestine (Dougal *et al.*, 2013). Given the greater numbers of bacteria in the hindgut and that many studies are based on faecal material, the following sections will focus on the organisms of the equine hindgut.

Region of gastrointestinal tract	**Total bacteria 10^6 /g (Kern *et al.*, 1974)**	**Total OTU's identified with a 99% similarity (Dougal *et al.*, 2013)**
Stomach	0.0001 - 0.0003	N/A
Ileum	0.000007	1980
Caecum	43.0	3747
Left ventral colon	N/A	3860
Terminal colon/ small colon	7.0	3731

Table 1.1 Bacterial numbers and OTU counts in different regions of the equine gastrointestinal tract.

1.3.1 Bacteria

In the horse large intestine, as in other mammals, the most dominant phylum has been largely reported as Firmicutes (Steelman *et al.*, 2012; Weese *et al.*, 2014; Hansen *et al.*, 2015; Zhao *et al.*, 2016; Clark *et al.*, 2018) with fewer studies identifying Bacteroidetes as the most abundant phylum (Beckers *et al.*, 2017; Salem *et al.*, 2018). However, Bacteroidetes is most widely reported as the second most abundant phylum, although, some studies have reported Verrucomicrobia as the second most abundant phylum (Steelman *et al.*, 2012; Costa *et al.*, 2015b; Beckers *et al.*, 2017). A more extensive list of studies with the proportions of Firmicutes, Bacteroidetes and Verrucomicrobia reported in each are listed in Table 1.2. Other major phyla are often variable in their order of abundance but include Fibrobacteres, Spirochaetes, Proteobacteria and Actinobacteria, as identified by the studies listed in Table 1.2. The discrepancy between studies of the most abundant phyla may be attributed to the horse populations sampled, the sampling medium (faeces or luminal contents) and region of the gastrointestinal tract sampled, DNA extraction kits as well as sequencing primers targeting the different hypervariable regions and the sequencing platforms (Pollock *et al.*, 2018).

The bacteria present in the equine hindgut can be divided into the following main functional groups: cellulolytic, amylolytic and the lactic acid utilisers. The organisms identified in the equine hindgut that belong to these groups will be discussed below.

The cellulolytic species are involved in the degradation of fibre. Between 3.9% and 12% of total anaerobic bacteria within the pony hindgut have been reported to be cellulolytic (Kern *et al.*, 1974; Lin & Stahl, 1995). Some of the main cellulolytic bacteria identified within the equine hindgut include *Fibrobacter, Ruminococcus and Clostridium*. Members of the *Fibrobacter* genus are characteristically non-spore forming, Gram-positive and are obligate anaerobes. *Fibrobacter* spp. Produce the VFAs succinic and acetic acid from the fermentation of xylan, glucose, cellulose and cellobiose (Lin & Stahl, 1995; Ransom-Jones *et al.*, 2012). Lin & Stahl, (1995) reported that 12% of the 16S rRNA genes isolated from equine caecal luminal contents were of the *Fibrobacter* genus and virtually all were identified as the species *F. succinogens*. In the colon, 4-5% of bacterial 16S rRNA again was identified as *F. succinogens*; however low levels of *F. intestinalis* were also detected. In contrast to total cellulolytic bacteria reported by Julliand et al. (1999) (3.8%), the figures reported in Lin and Stahl's study were much higher, up to 12%. However, the work by Lin and Stahl was performed on a single

pony and differences in the diets of the animals used in the two separate studies may account for these variations.

Ruminococcus spp. have also been identified with *R. flavefaciens* as the most prevalent species (Julliand *et al.*, 1999; Hastie *et al.*, 2008). In common with *Fibrobacter, Ruminococcus spp*. are Gram-negative, non-spore forming anaerobes which ferment cellulose and cellobiose (Van Gylswyk & Roche, 1970).

Other species of cellulolytic bacteria identified in the equine hindgut include members of the genus *Clostridium*. *Clostridium* is considered as a diverse genus consisting of commensal, fibre-degrading and pathogenic species and features heavily in healthy horses (Costa *et al.*, 2012; Steelman *et al.*, 2012). In total, 33 species of *Clostridium* were identified in equine faecal samples by Steelman *et al*., (2012).

Lactic acid producing and lactic acid utilising bacteria are able to tolerate conditions of a low pH and the presence of bile salts (Dicks & Botes, 2010). The acidity of the stomach and proximal region of the duodenum are prime regions of inhabitancy for these bacteria and are involved in initial carbohydrate digestion and providing a competitive environment against pathogenic species (Yuki *et al.*, 2000; St-Pierre *et al.*, 2013). However, in the hindgut when conditions become favourable, amylolytic species including *Streptococcus* and *Lactobacillus* can proliferate and rapidly become dominant causing hindgut disturbance which has been proposed to have systemic effects, including development of laminitis (section 1.4.3) (Milinovich *et al.*, 2008). The virulence and rapid establishment characteristics of lactobacilli have led to them being a target for use as a probiotic (section 1.4.10). Lactic acid producing bacteria interfere with the growth of other bacteria and can have inhibitory effects (Vandenbergh, 1993). When *Lactobacillus pentosus WE7* was administered orally to horses it was found to be inhibitory against various gut pathogenic bacteria including: *Salmonella* spp., *E. coli*, *Streptococcus zooepidemicus*, *C. difficile* and some inhibition against *C. perfringens* (Weese *et al.*, 2004).

Lactobacilli are anaerobic, non-spore forming, Gram-positive bacteria that ferment glucose to form lactate (Morita *et al.*, 2007). Lactobacilli sharing 84-98% identity were discovered in the following regions of the gastrointestinal tract: *L. salivarius* in the colon, *L. mucosae* in the caecal contents and *L. delbrueckii* and *L. salivarius* in the rectum (Al Jassim & Andrews, 2009). In healthy *ad lib* fed horses, *L. delbrueckii* and *L. mucosae* were identified in the stomach contents (Milinovich *et al.*, 2007).

Streptococcus bovis has been identified as an inhabitant of the healthy horse gut, but at lower levels than cellulolytic bacteria. *Ruminococcus flavefaciens* represented 5.85% in comparison to 2.67% of *S. bovis* of total bacteria recovered from caecal luminal contents of horses (Hastie *et al.*, 2008). The proliferation of *Streptococcus* has been linked with the onset of laminitis; however recent research has suggested that absolute abundance of *Streptococcus* spp. may not be as important relative to other changes during the onset of laminitis (Steelman *et al.*, 2012). This is reflected in values of 24.10% of total operational taxonomic units (OTUs) accounting for *Streptococcus* in healthy horses and a lower value of 16% *Streptococcus* OTUs in laminitic horses (Steelman *et al.,* 2012). However, it is worth noting in Steelman's paper, a figure of 74.88% of the bacteria accounted for *Streptococcus* in one individual, an unlikely occurrence that may have resulted in bias of the data.

Lactic acid utilising bacteria can ferment lactic acid into more useful VFAs that can be absorbed by the host (butanoic, acetic or propanoic acids). The genera of lactic acid utilising bacteria identified in the equine hindgut have been found to belong to *Selenomonas*, *Veilonella* and *Megasphaera* (Maczulak *et al.*, 1985; Daly & Shirazi-Beechey, 2003; Al Jassim *et al.*, 2005). Some species of bacteria have been identified to be both lactic acid producing and utilising (Al Jassim *et al.*, 2005; Ghali *et al.*, 2011). In normal conditions it is likely that lactic acid utilising bacteria and lactic acid producing bacteria are in balance to maintain homeostasis and are present in the regions of 10^5 and 10^6 CFU/ml in the caecum and colon (Fombelle *et al.*, 2003; Respondek *et al.*, 2008).

The bacterial populations that inhabit the equine hindgut are incredibly diverse. As reported by Daly *et al*., (2001) 89% of bacterial DNA sequences did not correlate to those previously found. More than a decade later in work by Shepherd *et al.*, (2012), a large proportion of unclassified bacterial sequences still remain - 38.1% of sequences isolated from faecal samples of healthy forage-fed horses were reported. Chimera checking and filtering was not carried out by Shepherd *et al*., and as suggested by Costa & Weese, (2012) the number of unclassified sequences might have actually been chimeras or artefacts. However in a more recent study where quality filtering was performed, 24% and 32% of sequences remained unclassified (Beckers *et al.*, 2017). Therefore, it can be concluded that information concerning the bacterial communities of the equine hindgut is still lacking.

Reference	Firmicutes (%)	Bacteroidetes (%)	Verrucomicrobia (%)
(Beckers *et al.*, 2017)	14.9	34.3	32.5
Steelman *et al.*, (2012) (control)	69.21	5.71	18.13
Steelman *et al*.,(2012) (laminitis)	56.72	9.94	27.63
Costa *et al.*, (2012) (healthy)	68.1	14.2	
Costa *et al*., (2012) (colitis)	30.3	40	
Zhao *et al.*, (2016) (Mongolian)	56	33	2
Zhao *et al.,* (2016) (Thoroughbred)	53	32	3
Liu *et al.*, (2014)(donkey male)	64	23	5
Liu *et al.,* (2014) (donkey female)	64	21	5
Costa *et al.*, (2015)*	36	0.9	27
Dougal *et al.*, (2013)*	48	40.4	
O'Donnell *et al.*, (2013)	48-74	17-31	2 to 11
Shepherd *et al.*, (2012)	43.7	3.7	4.1
Dougal *et al.*, (2017)	35	52	
Rodriguez *et al.*, (2015)(diarrhoea +)*	45	30	10
Rodriguez *et al*., (2015) (diarrhoea -)*	58	28	4
Ericsson *et al.*, (2016) (dorsal colon luminal contents)†	39	41	11
Peachey *et al.*, (2018)	34	39.9	12
Clark *et al.*, (2018)	49.9	34.5	
Salem *et al.*, (2018)	31.64	39.04	3.29
Proudman *et al.*, (2014)	53	42	0.002
Weese *et al.*, (2014) (pre-parturition)*	62	3.5	17
Weese *et al.,* (2014) (post-parturition)*	61	4.5	18
Weese *et al.*, (2014) (non-pregnant controls)*	63	3.5	12
Kristoffersen *et al.*, (2016)(caecal contents)*	65	22	4
Hansen *et al.*, (2005)(caecal contents)	65	20	
Whitfield-Cargile *et al.*, (2015) (foals at 3 weeks)	77	8.27	2.35
Whitfield-Cargile *et al.,* (2015) (foals at 5 weeks)	70.1	10.8	3.65
De Almeida *et al.,* (2016)	50.22		15.13

Table 1.2 The top three most widely reported phyla in equine faeces/hindgut microbiome studies. *Relative abundances extracted from bar graph †relative abundances calculated using raw data deposited in SRA (PRJNA322656) by Ericsson *et al.,* (2016). All values are reported means or ranges, when presented by authors. Table adapted from Beckers *et al.,* (2017).

1.3.2 Protozoa

Protozoa are present in the hindgut of the horse, but knowledge about their function and behaviour is limited. Protozoa are larger than bacteria but are present in lower counts and therefore are present at a similar biomass. The numbers of protozoa appear to be higher in the colon than in any other region of the hindgut (Dougal *et al.*, 2012). Detection of protozoa in the faeces of 5 day old foals indicated that establishment in the hindgut has begun by this age (Egan *et al.*, 2010). Protozoa detected in equine faeces may belong to the same order as ruminant protozoa, which includes Entodiniomorphida and Vestibuliferida. Early microscopic examination studies of equine protozoa revealed most are ciliated and are relatively similar to those found in the rumen (Becker, 1932). Species of ciliates appear to differ in terms of body style including size, shape, contractility and type of cilia present (Hsiung, 1930; Davis, 1941). These differences may indicate that species of protozoa are host specific (Snelling *et al.*, 2011). In a study by Kobayashi *et al*., (2006) a major species isolated from equine faeces was *Bundleia postciliata*, representing up to 54.7% of total protozoa. Other major species identified in the same study included *Didesmis quadrata* (up to 17.2%) and *Ditoxum funinucleum* (up to 15.8%). In horses grazing exclusively on pasture, 15 different genera of ciliates have been identified in equine faeces and *Spiridinium* and *Triadinium* were of the most abundant (Fernandes *et al.*, 2014). In horses fed grain based diets the following have been shown to be the most dominant genera: *Bundleia*, *Blepharocorys* and *Polymorphella* (Gürelli & Göçmen, 2011, 2012; Göçmen *et al.*, 2012).

It has been suggested that protozoa in the equine hindgut do not play an essential role in fermentation (Moore & Dehority, 1993). However later work by Laho *et al.*, (2013) demonstrated that protozoa isolated from zebra faeces were actively involved in the digestion of plant structural polysaccharides using α–amylase and inulinase. The zebra, also a member of the genus *Equus* and a hindgut fermenter, appears to share similar protozoal populations with the domestic horse (Laho *et al.*, 2013). Therefore, it is likely the protozoal populations of the horse behave in a similar way to the zebra. Furthermore, ciliates isolated from the equine caecum contain genes that code for cellulose and hemicellulose breakdown (Snelling, 2013). In the same work, ciliate genes were isolated from the equine caecum that proved to be homologous to genes in rumen-isolated ciliates (Snelling, 2013).

It is thought that protozoa may influence fermentation products, gut environment, bacterial population and health of the host. Complete defaunation (removal of protozoa from the

equine hindgut) resulted in an effect on the digestibility of the dry matter of a feed but not fibre digestibility (Moore & Dehority, 1993).

1.3.3 Fungi

Some fungal species found in the equine gastrointestinal tract may arise from dietary and environmental sources, although it is unknown whether these species are able to colonise in the gut and what impact they may have on the host (Doxey *et al.*, 1990). The other type of fungi found in the equine gastrointestinal tract, known as anaerobic fungi (Neocallimastigomycota), have been shown to contribute towards fibre degradation in ruminants (Gruninger *et al.*, 2014). Knowledge of anaerobic fungi in the equine hindgut is limited and has been largely extrapolated from ruminant studies (Edwards, 2019). Ruminant anaerobic fungi possess rhizoids which act to penetrate parts of plant material that bacteria and protozoa cannot. It is thought that fungi contribute towards the breakdown of cellulose and xylans in the gut and are particularly important for very high fibre diets (Krause *et al.*, 2013). When fungus-free sheep were dosed with *Neocallimastix* spp., the consumption of a straw-based diet increased by 40%, inferring the importance of fungi for the digestion of high fibre diets (Gordon & Phillips, 1993). Fungi account for approximately 10% of microbial organisms in ruminant animals, however this information and the extent of fungal contribution towards fibre degradation and overall energy requirements is unknown in the equine hindgut (Krause *et al.*, 2013). However, *P. citronii* isolated from equine faeces has been shown to produce formic and acetic acids, and ethanol as well as small amounts of fumaric, malic and succinic acids, suggesting a similar role of anaerobic fungi in the equine as in the ruminant (Julliand *et al.*, 1998). The rate of cellulose degradation was performed most rapidly and to the greatest extent by the donkey strain, which may indicate a greater adaptation of the donkey to a poorer quality diet (Julliand *et al.*, 1998). A small number of studies have identified the following genera of anaerobic fungi in the equine hindgut and faeces including: Neocallimastix, *Piromyces, Caecomyces, Orpinomyces, Anaeromyces* and the clades NG3 and AL1 (Gold *et al.*, 1988; Julliand *et al.*, 1998; Liggenstoffer *et al.*, 2010; Mura *et al.*, 2019). The distribution of anaerobic fungi in the latter regions of the hindgut has received preliminary investigation in one horse (Mura *et al.*, 2019). The latter study found that faeces were suitable for making a qualitative assessment of anaerobic fungi of the distal regions of the equine hindgut. For a quantitative assessment, faeces may be unsuitable as the relative abundances and concentrations of anaerobic fungi varied considerably between regions. Further work to establish the precise role of fungi in the equine gastrointestinal tract

is required to understand the contribution of fungi to the health of the host and digestion of plant fibre.

1.3.4 Archaea

Of the microorganisms found in the equine hindgut, knowledge of archaea populations is the most limited. In the equine gut, archaea represent around 3.5% of the total microbial cell numbers (Yamano *et al.*, 2008). The archaea population include the methanogens. Methanogenic archaea convert waste products of fermentation (carbon dioxide and hydrogen) into methane which is then expelled from the tract (Hook *et al.*, 2010). This promotes the growth of other bacterial species that are more efficient at fermentation. Organisms showing similarity to the rumen species *Methanobrevibacter ruminantium* and *Methanocorpusculum parvum* have been identified in equine faeces (Yamano *et al.*, 2008). However in work by Lwin & Matsui (2014) a large proportion of sequences showed low similarity to already identified methanogens, suggesting that novel species may exist within the equine hindgut. It is likely that methanogenic organisms are in low abundance in comparison to ruminants, as methane production is estimated to be 3-4 times smaller per kg of feed in horses compared to ruminants (Vermorel, 1997). In work by Fernandes *et al.*, (2014) a shift in bacteria was observed in response to dietary change; in contrast, archaea remained stable in equine faeces. Rather than relying on components of the host diet, archaeal organisms source their nutrition from hydrogen produced as a by-product of fermentation by the populations of bacteria in the hindgut.

1.4 Factors correlated with changes in the equine microbiome and metabolome

Emerging evidence suggests that multiple factors have an influence on the composition of the equine hindgut microbiome. It is important to determine factors associated with changes in the hindgut microbiome (and the functional metabolome) to provide knowledge of factors that need to remain constant in experimental trials studying the microbiome. In addition, investigation of how management factors such as change in diet, which are known to be associated with altered risk of colic, are important. These enable us to further understand how the microbiome and associated metabolome function in equine health and disease, with important consequences for equine welfare.

1.4.1 Age

The development of the hindgut microbiota in herbivores is essential for the breakdown of dietary plant fibre to meet the energy requirements of the animal. Studies have shown that the foal hindgut microbiota reaches adult maturity by four weeks of age (Faubladier *et al.*, 2014). Sources of bacteria are thought to be maternally and environmentally derived (Quercia *et al.*, 2019). Aero-anaerobic bacteria including Lactobacilli, Enterococci and Enterobacteria colonise within the first few days of life (Julliand *et al.*, 1996). Cellulolytic bacterial species begin to establish during the first week when the foal begins to consume fibre and perform coprophagy (Faubladier *et al.*, 2014; Quercia *et al.*, 2019). However, post-weaning an increase in Fibrobacteres (cellulolytic species) increased in foals, likely because of the transition from a milk based to a non-milk based diet (Costa *et al.*, 2016).

In the elderly horse an alteration in the gut microbiota has been shown. A reduction in bacterial diversity was seen in a group of horses aged 19-28 years in comparison to a group aged between 5 and 12 years (Dougal *et al.*, 2014). Reduced gut bacterial diversity is also a characteristic in the aging human (Hopkins & Macfarlane, 2002). Increased digesta transit time, a lower requirement for dietary energy or less efficient mastication associated with poor dentition may be responsible for a reduction in bacterial diversity in the aging horse (Lowder & Mueller, 1998; Morley, 2007).

1.4.2 Sex

Differences in bacterial groups in colon and faecal samples between male and female macaques has been demonstrated (McKenna *et al.*, 2008). However it has been found in mice that sex is less of a factor influencing microbiota composition than genotype (Kovacs *et al.*, 2011). The use of culturing techniques demonstrated a significant difference between stallions and mares in Maiduguri, Nigeria (Mshelia *et al.*, 2018). In a French cohort of endurance horses there were no significant differences in the faecal microbiota between sexes (male, female and gelding) (Massacci *et al.*, 2020). Therefore, sex may have some effect on the gut microbiota composition, but other factors may override this.

1.4.3 Diet

The effect of diet on the equine hindgut and faecal bacteria is a well-researched area and is regarded as an important influencing factor (Willing *et al.*, 2009; Daly *et al.*, 2012; Van den Berg *et al.*, 2013; Fernandes *et al.*, 2014). Resistant starch, present in substantial quantities in grain and as storage carbohydrates (fructan) in plants cannot be broken down in the small

intestine (Longland & Byrd, 2006; National Research Council, 2007). Therefore, undigested material passes through into the hindgut where it is rapidly fermented by the anaerobic bacteria of the hindgut producing high levels of lactic acid. This causes the environmental pH to drop and is known as lactic acidosis (Garner *et al.*, 1977). In normal conditions a balance between lactic acid producing bacteria and lactic acid utilising bacteria prevents acidosis. However in times of carbohydrate overload the lactic acid utilising bacteria are no longer able to process large quantities of lactic acid allowing *Lactobacillus* and *Streptococcus* species to dominate (Milinovich *et al.*, 2008; Respondek *et al.*, 2008).

In experimental studies, the infusion of oligofructose which is indigestible pre-caecally has been shown to have profound effects on the equine hindgut microflora. Prior to oligofructose administration, the caecal fluid of horses consisted largely of Gram-negative species with 20% Gram-positive species of bacteria. Within 4-6 hours post-oligofructose administration, caecal fluid contained more than 50% Gram-positive organisms. The horses investigated also developed profuse diarrhoea and a drop in caecal pH, both indicators of gastrointestinal disturbance. Concurrently laminitis signs developed including lameness and a shifting of weight from foot to foot supporting a dietary related role in some forms of laminitis (Milinovich *et al.*, 2007).

Abrupt changes in diet from predominantly forage to concentrate feeding has been shown to result in higher counts of amylolytic bacteria including *Streptococcus* and *Lactobacillus* species (Van den Berg *et al.*, 2013). In one study, within 29 hours of barley being introduced to a hay diet, an increase in *Streptococcus* and *Lactobacillus* were observed in the colon (de Fombelle *et al.*, 2001). In a separate study an increase in Streptococcus in the caecum was seen 14 hours after introducing barley to the diet (Kristoffersen *et al.*, 2016). The work by de Fombelle *et al.,* (2001) also identified an increase in lactic acid levels and a decrease in the acetic and butanoic acid: propanoic acid ratio. This is consistent with work by Willing *et al.,* (2009) where horses fed on a concentrate diet had 10 times the number of lactic acid producing bacteria compared to horses fed a forage only diet. In the concentrate diet group, 73% of sequences were Gram-positive bacteria, whereas in the forage-only diet 46% of sequences were Gram-positive. In comparison, in a group of grass fed horses that were sampled post mortem at an abattoir, only 3% of sequences were matched to lactic acid producing bacteria (*S. bovis* and *L. salivarius*) (Daly *et al.*, 2001). Kristoffersen *et al.,* (2016) demonstrated through shotgun metagenomics sequencing that bacterial genotypes associated with a high starch diet were carbohydrate metabolism related, whereas genotypes for the hay diet were associated with monosaccharides and glycosidic hydrolyses.

The genotype findings correlated with levels of SCFA, which were elevated when horses were fed on a high starch diet. Longitudinal studies have been performed to investigate temporal stability of the microbiota in forage fed diets, but this has not been performed in concentrate and forage fed diets and is an area for future study.

The switching of forage types appears to induce changes to the hindgut microbiota at a slower rate compared to abrupt introductions of concentrates. Following an abrupt transition from stabling (with a hay based diet) to pasture on the same day, changes in relative abundances of bacterial communities were seen by day four (Fernandes *et al.*, 2014). The latter study concluded a highly diverse bacterial community may be an evolutionary adaptation of herbivore to rapidly respond to forage availability and quality.

A dietary change from feeding hay to silage or haylage (forages that have been fermented and preserved by acidification) has been found to be associated with an increase in *Streptococcus* and *Lactobacillus* (Muhonen *et al.*, 2009). In addition forage-fed diets supplemented with either oil or starch showed that the majority of phylum differences between groups were in Firmicutes followed by Bacteroidetes, Proteobacteria, Actinobacteria and Spirochaetes (Dougal *et al.*, 2014). The largest proportion of operational taxonomic unit (OTU) sequences shared between horses was when a forage diet was fed (15.9%) followed by the forage and oil (10.3%) and the least shared between the forage and starch diet (5.4%) (Dougal *et al.*, 2014). Work by Hansen *et al.,* (2015) has shown that feeding a forage diet with oats reduces caecal bacterial diversity. In the latter study, the phylum Tenericutes were less frequent in samples from horses fed an oat-based diet and increased levels of VFAs were also observed. From these studies it can be concluded that diets rich in readily fermentable substrate (e.g. starch) produce excess VFAs and lactic acid that may cause the microbiota to become more variable and hence less stable as there is a drop in pH and the environment becomes less favourable to cellulolytic species. Although bacterial change and increased levels of VFAs were observed in Hansen's study, no clinical signs of lactic acidosis (e.g. diarrhoea or associated laminitis) were reported; therefore, the changes were considered to be below the threshold for causing clinical disease. Studies investigating the change in microbiota over time when ponies were maintained on high fibre diets showed a high degree of stability (Blackmore *et al.*, 2013; Dougal *et al.*, 2017). After 6 weeks on a high fibre diet, 65% of bacteria were retained (Dougal *et al.*, 2017), which is a similar figure found in a cohort of humans on mixed diets after one year (70%) (Faith *et al.*, 2013).

The frequency of meals fed can also have an effect on the amount of lactate and VFA (indicators of microbial activity) present in the hindgut. Respondek *et al.,* (2008) fed 237g of starch per 100kg of bodyweight spread over three meals throughout the day, whereas Medina *et al.,* (2002) fed 136g starch per 100kg of bodyweight spread over two meals throughout the day. Interestingly, the levels of lactic acid and VFA were lower in Respondek's study of higher starch but higher meal frequency. Feeding one large concentrate meal rather than smaller meals throughout the day led to wider fluctuations in bacterial abundance, particularly in *Streptococcus,* a bacterium related to lactic acid production and dysbiosis (Venable *et al.*, 2017b). These studies indicate that feeding concentrate in multiple smaller meals is less likely to result in gut dysbiosis than feeding as one large meal.

1.4.4 Season

Season is one area that has been investigated but may relate more to dietary change according to ambient weather conditions rather than the effect of season (e.g. temperature, humidity) itself. In a study by Kobayashi *et al*., (2006) involving Hokkaido native horses and light horses who were raised together on the same pasture following birth. During the summer, an increase in protozoal counts was observed in the native horse group. This seasonal change was considered likely to be linked with the fibrous material consumed by the horses at that time of year. Individual plant selection was considered to be a potential explanation as to why little increase was seen in protozoal counts in the light horse group. In more recent work by Salem *et al.,* (2018) a group of horses kept solely at pasture were studied in the UK over a 12-month period. Haylage was introduced to the diet over the winter and an increase in Gram-negative bacteria (specifically Fibrobacters), was observed, which agrees with the work by Koike *et al*., (2000). Furthermore, Salem *et al.,* (2018) observed a relative increase in Spirochaetes and a decrease in Firmicutes after the introduction of haylage to the diet. In a multivariable model, ambient weather conditions were associated with changes in microbiota. It was not possible to determine if this was an effect of weather on grass composition or potential changes in the soil microbiota (Salem *et al.*, 2018). This is an area that warrants further study in relation to the equine gut microbiome.

1.4.5 Host genetics and breed

In a Latin square study design denaturing gradient gel electrophoresis revealed different banding patterns of DNA extracted from faeces of different breeds of dog, with no effect of diet (Simpson *et al.*, 2002). In horses, cellulolytic species of bacteria, in particular *F. succinogens* appear to be more dominant in the flora of native types of horses than light

horses (Koike *et al.*, 2000). In a comparison of the hindgut microbiota of Thoroughbred and Mongolian horses, it was observed that Proteobacteria (*Salmonella*, *Helicobacter pylori*) and *Streptococcus* were significantly higher in the Thoroughbreds than the Mongolian horses (Zhao *et al.*, 2016). The two groups of horses in the work by Zhao *et al.,* (2016) were sampled from the management conditions typical to that breed – stabled Thoroughbreds and prairie grazed Mongolian horses, therefore it cannot be determined whether the differences observed were environmental, dietary or genetic. Furthermore native Hokkaido horses have been demonstrated to have a more diverse microbial population than their light horse counterparts, even when born and raised in the same conditions (Yamano *et al.*, 2008). This suggests possible differences in genetic make-up between light and native Hokkaido horses, which may influence the presence of bacterial species. Furthermore, thoroughbred horses and Japanese native ponies maintained in the same conditions on the same premises were also found to have a different composition of methanogens within the hindgut (Lwin & Matsui, 2014). Therefore, as well as bacteria other microorganisms in the equine hindgut may be determined by host genetics or behaviour. However, gastrointestinal anatomy and function as well as foraging behavioural traits may differ between the native and light Hokkaido horses making it difficult to interpret how strongly the gut microbiota is affected by genetics in horses. More recently, in a larger cohort of horses (n=189) managed in similar conditions subtle, but significant differences in beta diversity were observed between Lusitano, Anglo Arabian and central European breeds (Massacci *et al.*, 2020). However, microbiota changes were not associated with genetic ancestry (Massacci *et al.*, 2020)

1.4.6 Body condition

Studies performed in mice provide evidence that the gut microbiome may influence the deposition of body fat. Microbiota from dietary-induced obese mice was transplanted into germ-free mice by Turnbaugh *et al.,* (2008). The germ-free mice that received microbiota from the obese mice were found to have greater fat deposition than those who received a microbiota transplant from lean donors. On culturing the microbiota from the dietary-induced obese mice it was found that an uncultured clade of the class Mollicutes, was dominant, which is normally of low abundance in the lean mouse. Bacterial end products including acetic acid, were also of higher concentration in mice with a greater abundance of Mollicutes (Turnbaugh *et al.*, 2008). In other species, including cats and dogs the microbiota of obese and lean groups were found to significantly differ (Handl *et al.*, 2013; Kieler *et al.*, 2016). In the horse, real-time polymerase chain reaction (PCR) revealed no significant differences in Firmicutes, Bacteroidetes or two specific cellulolytic bacteria (*F. succinogens*

and *R. flavefaciens)* between five obese and five moderate-condition mares (Shepherd *et al.*, 2014). Faecal changes in propanoic, butanoic and acetic acids were also not seen between the two groups of mares. The physiological differences between moderate-condition and obese may not have been sufficient to observe differences between the two groups in the latter study. In addition, the use of in-depth bacterial sequencing techniques and metabolic profiling may provide more information.

1.4.7 Stress, transport and exercise

Transportation is often associated with stress-related behaviours in horses and the risk of colic is increased if a horse has travelled recently (Hillyer *et al.*, 2002; Tateo *et al.*, 2012). Stress has been shown to alter the microbiota in mice (Gao *et al.*, 2018). For horses, in addition to stress, transportation in itself can be a marker of management change including the removal of water or feed for several hours. Lactobacilli and Streptococci have been found to be increased in abundance in caecal contents of cannulated horses during transport (Perry *et al.*, 2018). However, a cross-over study design on this small sample size (n=3 per group) would have enabled more robust conclusions of the effect of transport on the caecal microbiota to be made. Furthermore, transport related changes have to be considered carefully as it is difficult to relate changes to direct effects of transport (e.g. physiological stress) or this being a marker for other factors such as dietary change.

Weaning (normally around 6 months of age) is a period of dietary change and physiological stress (separation from the mare). According to work by Mach *et al.,* (2017) the method of maternal separation influenced the gut microbiota. The genera *Eubacterium*, *Coprococcus*, *Clostridium XI*, and *Blautia spp*. were negatively correlated with salivary cortisol levels (an indicator of stress) but positively associated with butanoic acid levels (Mach *et al.*, 2017). The group of foals that underwent progressive weaning (gradual maternal separation over 4 weeks) had higher levels of anaerobic fungi (indicating a gut microbiota better prepared for fibre breakdown) than an abrupt weaning (outright separation from day 0) method. There is no current evidence about whether the microbiota alterations that occur at weaning have a long-term impact on the health of the horse.

There is evidence that exercise can have an impact on the gut microbiota (Mach & Fuster-Botella, 2017). However, minimal effects of exercise have been described in the horse (De Almeida *et al.*, 2016). A recent change in exercise program has been linked with colic (Hillyer *et al.*, 2002). However the link between exercise and colic may be indirect, with the changes that often coincide with alterations in exercise program e.g. diet also playing a role (Hillyer

et al., 2002). The direct involvement of the microbiota in exercise-associated colic has not been investigated.

1.4.8 Inter-individual variation

It has emerged that an inter-individual variation exists between hindgut bacterial communities of horses. In other species, including the dog and the macaque, differences in microbiota of individuals have been shown (Simpson *et al.*, 2002; McKenna *et al.*, 2008; Handl *et al.*, 2013). This is apparent in horses even when the individuals have been born and raised in the same conditions (Kobayashi *et al.*, 2006) or maintained on the same dietary regime for sustained periods (Willing *et al.*, 2009; Blackmore *et al.*, 2013; Dougal *et al.*, 2017). Furthermore individual clustering by horse has also been observed when characterising microflora of the stomach, suggesting inter-horse variation in bacteria is common to other areas of the gastrointestinal tract (Perkins *et al.*, 2012). A strong trend for individuality, even when environmental conditions are consistent, suggests that there are many factors which influence the intestinal microbiota.

The possibility of the horse possessing a core microbiome has been explored in a group of ten horses (Dougal *et al.*, 2013). It was found that 5-15% of caecal OTUs were shared between individuals. In the faeces of healthy human subjects, 40% of phylotypes were consistent between individuals (Jalanka-Tuovinen *et al.*, 2011). This work suggests in comparison to other species the core microbiome shared between horses is small and a substantial proportion of it is readily interchangeable; given this evidence it is unsurprising that there is such variation between horses.

The effect of inter-individuality on the hindgut metabolome has received less attention than the microbiome. Horses consuming the same diet have been found to have different individual fermentation rates and the inter-individual variation was much greater than in any of the other fibre fermenting species studied (Uden & Van Soest, 1982). In later work, Blackmore *et al.,* (2013) maintained a group of ponies on the same diet for 11 weeks it was found that in general VFA concentration increased in all animals; however the levels of specific major VFAs (acetic, butanoic, propanoic and valeric acids) varied between individuals.

1.4.9 Antimicrobials

Antibiotics are frequently used in equine veterinary medicine to treat a variety of bacterial infections but have been shown to also have effects on the gut microbiota. (Ferran *et al.*,

2013; Costa *et al.*, 2015b). In mares given a five-day course of intramuscular (procaine penicillin and ceftiofur sodium) and oral (trimethoprim sulfadiazine) antimicrobials changes were seen at phylum level, demonstrating the extent of which antimicrobials can alter the gut microbiota (Costa *et al.*, 2015b). Although changes in the microbiota were observed for both intramuscular and oral treatments, the greatest impact was observed after trimethoprim sulfadiazine. Post-treatment (25 days) microbiota populations resembled those of pre-treatment; however some changes were still evident (Costa *et al.*, 2015b). It is important to understand the effects of antimicrobial treatment on commensal gut bacteria, as changes to these populations may create an opportunity for other pathogenic bacteria to invade (Cotter *et al.*, 2012). Acute colitis, frequently caused by the bacterium *Clostridioides difficile* in the horse is a reported complication of antimicrobial treatment (Baverud *et al.*, 1997).

1.4.10 Probiotics

Probiotics are live microorganisms (given as a feed supplement) with the aim of providing benefits to the host in addition to providing nutrition (Weese, 2003). Because of the sensitivity of the equine microbiota to change, probiotics have the potential to modify the intestinal bacteria for the prevention and treatment of disease. A number of bacterial genera and yeast species have been evaluated for use in horses, including: *Saccharomyces*, *Lactobacillus*, *Enterococcus* and *Streptococcus* (Schoster *et al.*, 2014). In particular some species of lactobacilli (*L. equi*, *L. crispatus*, *L. johnsonii*, *L. reuteri* and *L. salivarius*) have shown to have positive effects for the equine host by preventing diarrhoea (Yuyama *et al.*, 2004). The mechanisms of lactobacilli to prevent pathogen invasion is by competing with epithelial binding sites as well as producing antimicrobial agents including 2,3-butanedione, hydrogen peroxide and carbon dioxide (Montes & Pugh, 1993; Naidu *et al.*, 1999). The effects of probiotic supplementation on the hindgut microbiota as a whole has received little attention, however probiotics may stabilise faecal pH by producing carboxylic acids (Ishizaka *et al.*, 2014). More work is needed in this area of research.

1.5 Gastrointestinal disease related to the hindgut microbiota

1.5.1. Colic

Colic is an umbrella term for abdominal pain, with many different underlying causes (Curtis *et al.*, 2019). Disturbance to the intestinal microbiota has been proposed as a potential contributing factor to some forms of colic including simple colonic obstruction and distension

(SCOD) and large colon volvulus (LCV) (Daly *et al.*, 2012; Weese *et al.*, 2014). In horses diagnosed with SCOD significantly fewer cellulolytic species including *Fibrobacter* and *Ruminococcus* spp. were present in the colonic contents than healthy grass-kept horses (Daly *et al*., 2012). In the latter study, the horses with SCOD shared a microbiome similar in structure to horses fed a concentrate diet. The authors concluded that the feeding of concentrates and SCOD both resulted in a less stable microflora. Turner *et al*., (2013) reported higher concentrations of acetone and methanol in the faeces of horses presented with SCOD than horses presented with non-gastrointestinal related disorders. One study was able to detect changes in the faecal microbiome preceding clinical presentation of a colic episode (LCV) in broodmares (Weese *et al.*, 2014). The pre-colic faecal microbiome was characterised by an increase in relative abundance of Proteobacteria and a decrease in Firmicutes.

Equine grass sickness (EGS) is a neurodegenerative disorder affecting the enteric nervous system (equine dysautonomia). Numerous studies have found a link between EGS and the presence of the bacterium *Clostridium botulinum* in the intestine (Hunter *et al.*, 1999; Garret *et al.*, 2002; Wagget *et al.*, 2010) and this theory has been supported by epidemiological studies (Wylie *et al.*, 2016). Work carried out to determine whether EGS has an effect on the wider microbiome has shown an increase in Bacteroidetes and a decrease in Firmicutes bacteria in horses with EGS compared to control groups (Leng *et al.*, 2015). Furthermore, the genera *Desulphovibrio* and *Veillonella* and the species *Veillonella parvula* increased in abundance. Whether a change in the gut microbiome is a cause or a consequence of EGS (due to effects on gut motility) is currently unknown.

Colitis (inflammation of the caecum and colon) has been associated with a number of bacterial pathogens including: *Clostridioides difficile*, *C. perfrigens* and *Salmonella spp*. (Weese *et al.*, 2001; Chapman, 2009). Costa *et al*., (2012) compared the faecal microbiota of healthy horses with those suffering from undifferentiated colitis. It was found that healthy horses had significantly higher abundances of Actinobacteria and Spirochaetes whereas horses suffering from colitis had a greater abundance of Fusobacteria, indicating that colitis may be a disease of gut dysbiosis rather than simply the overgrowth of a single pathogen.

High burdens of gastrointestinal parasite infections by strongyles or cestodes (tapeworms) are known risk factors for colic (Archer and Proudman 2006). Different types of strongyles have various pathogenic effects on the equine gastrointestinal tract resulting in colic due to emboli formation and segmental intestinal necrosis (*S. vulgaris*), small intestinal obstruction

(*Parascaris equorum*) or weight loss, diarrhoea and gut intussusception (cyathostomins) (Proudman, 2017). Epidemiological studies have shown that the equine tapeworm (*Anoplocephala perfoliata*) is associated with increased risk of spasmodic colic and colic due to ileal impaction (Proudman *et al.*, 1998; Little & Blikslager, 2002).

1.5.2 Gastrointestinal parasites

Horses are frequently born intestinal parasite free, but transmission of Strongloides can occur in utero (Gelberg, 2017). Foals usually become infected within the first week of life and will retain some level of infection for the rest of their existence. Intestinal parasites share the gut environment with the intestinal microbiota but despite this little is known about how they interact. Various parasitic infections have been shown to alter intestinal bacterial diversity in humans (Lee *et al.*, 2014), rabbits and rodents (Walk *et al.*, 2010; Cattadori *et al.*, 2016), but in other species including horses (Clark *et al.*, 2018; Peachey *et al.*, 2018), goats (Li *et al.*, 2016), cattle (Li *et al.*, 2011b) dogs and cats (Šlapeta *et al.*, 2015) parasite infection has not altered bacterial diversity. However, in young horses (1 year old) bacterial richness was reduced in those with high strongyle burdens compared to those with low burdens (Peachey *et al.*, 2019). Age may be a factor in determining the extent of the effect of parasitic infection on the gut microbiota. In pigs, 13% of genera detected in the proximal colon were altered in response to *Trichuris suis*, in particular *Fibrobacter* and *Ruminococcus* reduced in abundance (Wu *et al.*, 2012). Tapeworm (*Hymenolepis* spp.), present in the small intestine of mice have been associated with an increase in Bacteroidetes in the stomach (Kreisinger *et al.*, 2015). This is an indication that the presence of parasites in the gastrointestinal tract can influence bacterial homeostasis. Also in murine studies, experimentally administered *Lactobacillus taiwanensis* increased regulatory T cells and helminth establishment, indicating a causal relationship in which commensal bacteria promote infection with an intestinal parasite (Reynolds *et al.*, 2014a).

Helminths have been found to stimulate the immune system both positively and negatively by possible interaction with the gut microbiota (Resende Co *et al.*, 2007; Maizels *et al.*, 2009). Murine studies have demonstrated that helminth-microbiota interactions have been found to be protective against allergies; currently known mechanisms of these interaction are detailed in a recent review (Brosschot & Reynolds, 2018). However, gastrointestinal parasitic infection itself can be damaging to the host resulting in signs of weight loss, damage to gut mucosa, migration and damage to other tissues. In a rat model, infection with a non-pathogenic tapeworm species (*Hymenolepis diminuta*) induced minor changes to the faecal

microbiome, but still elicited an immune response (Parfrey *et al.*, 2017). In mice, transfer of a helminth-modified microbiome (without the live infection) reduced airway inflammation in an allergic asthma model (Zaiss *et al.*, 2015). Therefore, non-disease-causing parasitic infections or a helminth-modified microbiome (in absence of helminths) may have future implications for immunomodulation therapies. To date, the impact of parasitic infection on host immune stimulation (Pavone *et al.*, 2011; Pittaway *et al.*, 2014) and microbiota-immunity correlations have received little attention in the horse (Clark *et al.*, 2018).

The composition of the microbiota may influence susceptibility of a host to parasitic infection. This was demonstrated in a recent study, where human subjects with more than 2.5 mM of indole as well as greater abundances of indole producing bacteria (*E. coli*, *Clostridium* spp. *Bacillus* spp.) were more likely to become infected with *Cryptosporidium* (Chappell *et al.*, 2016). In the horse, Clark *et al.*, (2018) identified ponies as either susceptible or resistant to strongyle infection. Susceptible ponies had lower abundances of commensal bacteria including *Ruminococcus, Clostridium* XIVa and members of the Lachnospiraceae family than resistant ponies. Furthermore, resistant ponies had lower faecal egg counts (FECs) and higher circulating eosinophils. The work demonstrated differences in the microbiota and immune status between ponies resistant and susceptible to strongyle infection under a typical equine husbandry regimen. However, because of uncontrolled factors (level of infection that the ponies were exposed to and differences in grazing behaviour) it is difficult to determine the mechanisms by which these differences occurred. Interestingly, the microbiota composition of susceptible and resistant ponies was similar at the start of the trial. Therefore, the post-infection microbiota alterations observed between the two groups of ponies might have been an indirect consequence of the reason for strongyle infection (e.g. grazing behaviour or host immunity) or as a response to infection. Further controlled studies investigating parasite-microbiota and host interactions in the horse are required to better understand susceptibility to infection.

The evidence discussed here suggests that interactions occur between the gut microbiota, the products they produce and parasites. A small number of studies have addressed the effect of strongyle spp. on the equine faecal microbiota, yet the effects of other major equine gastrointestinal parasites including tapeworm are still to be investigated (Clark *et al.*, 2018; Peachey *et al.*, 2018; 2019). Furthermore, there are no published findings of the effect of other major parasites including tapeworm on the gut microbiome and metabolome.

1.6 Introduction to tapeworm infection

Anoplocephalid cestodes (tapeworm) are gastrointestinal parasites that affect horses worldwide. The three species that infect horses are *Anoplocephala perfoliata* (Goeze, 1782) *Anoplocephala magna* (Abildgaard, 1789) and *Anoplocephaloides mamillana* (Mehlis, 1813). *A*. *perfoliata* will be the focus of this thesis as it is the most prevalent and most pathogenic species and therefore the species of greatest clinical importance (Owen *et al.*, 1989; Chapman *et al.*, 2002).

1.6.1 Morphology and lifecycle

The mature tapeworm *A. perfoliata* measures between 2 and 8 cm in length (Proudman, (1994). It has a flattened body of which proglottids (segments) develop from the scolex downwards. Four 'suckers' are located on the scolex, which allow the tapeworm to attach to the gut mucosa and nutrients are absorbed via the cuticle. Some images of *A. perfoliata* at immature and mature stages (gravid proglottids) are shown in Figure 1.2a and b, respectively.

Mature tapeworms attach to the mucosa of the intestinal tract and produce gravid proglottids which contain eggs. On the process of passing through the intestinal tract, eggs are released from the segments and pass out with the faeces. The oribatid mite is the intermediate host for *A. perfoliata* and is free-living in pasture. As well as acting as an intermediate host for *A. perfoliata,* the oribatid mite plays an important role in the decomposition of organic matter in the soil (Norton & Behan-Pelletier, 2009). Per $1m^2$ of pasture, oribatid mites are present in numbers of a few thousand to hundreds of thousands (Tomczuk *et al.*, 2015). On ingestion of tapeworm eggs by the oribatid mite, the eggs develop into the invasive form – cysticercoid – over a 3-month period. The oribatid mite is then taken in by the grazing animal which leads to the establishment of infection. A diagram of the indirect lifecycle of *A. perfoliata* is shown in Figure 1.2c.

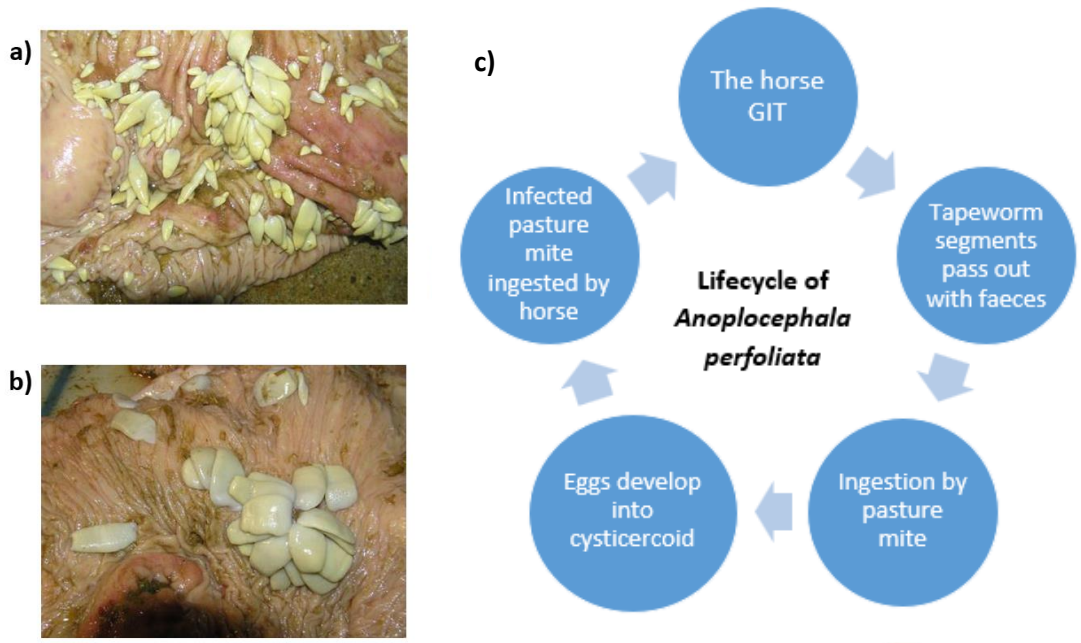


Figure 1.2 Images of *Anoplocephala perfoliata* and a schematic diagram of the lifecycle. The images of *A. perfoliata* were taken from (Tomczuk *et al.*, 2014) (open access, link to licence: https://creativecommons.org/licenses/by/3.0/legalcode) and are shown in **a)** immature stages and **b)** mature stages (gravid proglottids). The parasite lifecycle, shown in **c)** was adapted from (Proudman, 2003)**.**

1.6.2 Prevalence

Tapeworm infection has been found to occur in horses from 5 months of age, but infection intensity is at its strongest in young horses (aged 0.5 - 5 years) and older horses (aged 15 years plus) (Proudman, 1994; Matthews *et al.*, 2003; Engell-Sørensen *et al.*, 2018). Female horses have been found to be at greater risk of *A. perfoliata* infection than entire males and geldings when sampled over one year from abattoir material (Nilsson *et al.*, 1995). However others have found no association between sex and *A. perfoliata* infection (Rehbein *et al.*, 2013; Lyons *et al.*, 2018). Lyons *at al.,* (2018) found Thoroughbreds had significantly higher number of tapeworms than non-Thoroughbreds, however other management factors were not controlled therefore it is difficult to conclude if this difference was truly breed related.

In abattoir studies, *A. perfoliata* was present in 7.4 - 65.4% of horses examined (Pavone *et al.*, 2011; Tomczuk *et al.*, 2014). The number of worms found in the intestinal tract ranged from 1-2,069 with mean burdens of 35-265 worms, (Nilsson *et al.*, 1995; Williamson *et al.*, 1997; Tomczuk *et al.*, 2014). Infection intensities of more than 100 worms have been reported in 8 - 39% of horses with *A. perfoliata* (Meana *et al.*, 1998; Rehbein *et al.*, 2013; Tomczuk *et al.*, 2014).

Season appears to have an influence over the number of worms recovered from the gastrointestinal tract. A greater presence of tapeworms during the autumn and winter than in spring and summer months has been reported in numerous studies (Nilsson *et al.*, 1995; Meana *et al.*, 2005; Rehbein *et al.*, 2013). It is noteworthy that these studies were performed in climates with cold winters and dry, warm summers. Studies carried out in maritime climates found a less prominent effect of seasonality – egg-producing tapeworms were present all year round (Agneessens *et al.*, 1998; Engell-Sørensen *et al.*, 2018). Moist and cool environments are more favourable conditions for the intermediate host (oribatid mite) and therefore the maritime climate may be more suited to this organism (Gergócs & Hufnagel, 2009). Engell-Sørensen and colleagues reported a slightly higher prevalence in the autumn, possibly because of a gradual increase in infection intensity over the grazing season. During the winter horses are more likely to be stabled and therefore protected from further infestation. In support of this others have found a positive association with and time spent on pasture and level of infection. (Trotz-Williams *et al.*, 2008).

1.6.3 Pathogenesis

Post-mortem examination of fifty tapeworm-infected horses in one study revealed that 17% of tapeworms were found at the ileo-caecal junction, 81% on the caecal wall, 1.7% in the terminal ileum and 0.2% in the ventral colon (Williamson *et al.*, 1997). Horses with low burdens of tapeworm (1-20) show little pathology and the infection is regarded to have little impact on health (Pearson *et al.*, 1993). According to Reinemeyer & Nielsen (2009), the impact of an average tapeworm infection on the horse is not known and effects are likely to be sub-clinical.

Inflammatory changes in the gut wall increase linearly with the number of worms present (Pearson *et al.*, 1993; Nilsson *et al.*, 1995). An increase in eosinophils and a decrease in lymphocytes at parasite attachment sites have been observed (Pearson *et al.*, 1993; Williamson *et al.*, 1997; Pavone *et al.*, 2011). Antibodies IgG(T) and IgE have been detected in the mucosal tissue at attachment sites, but the host response to *A. perfoliata* is not yet fully understood (Pittaway *et al.*, 2014). Lesions characteristic of *A. perfoliata* infection included: mucosal ulceration, inflammation and thickening, most severely when high numbers (more than 100) *A. perfoliata* were present (Pearson *et al.*, 1993). Attachment of *A. perfoliata* to the mucosa of the ileo-caecal junction may change the diameter of the gut. In severe cases prolapse of the terminal ileum into the ileo-caecal junction (intussusception) has been reported (Williamson *et al.*, 1997). The risk of colic also increases with increasing

infection intensity (Proudman *et al.*, 1998). The types of colic associated with tapeworm infection are caecal intussusceptions, caecal rupture, spasmodic colic and ileal impaction colic (Owen *et al.*, 1989; Proudman *et al.*, 1998; Proudman & Holdstock, 2000; Ryu *et al.*, 2001; Boswinkel & Sloet van Oldruitenborgh-Oosterbann, 2007). Many of these require surgical intervention and if left untreated can result in the death of the animal.

The precise mechanisms by which tapeworm-associated colic develops are not known but several theories have been presented of how the parasite may alter intestinal function including normal gut motility. Damage by *A. perfoliata* to the enteric nervous system has been reported, but the extent of this is yet to be clarified (Pavone *et al.*, 2011). It has been speculated that the parasite attached to the mucosa may alter the host parasympathetic nerve transmission by the release of acetylcholinesterase, although work has not been carried out to confirm this theory (Lee & Tatchell, 1964). It has also been suggested by Pavone and colleagues that the damage caused to the intestinal mucosa by the parasite attachment (ganglionitis) is similar to that presented in inflammatory bowel diseases, resulting in impaired motility and transit (De Giorgio *et al.*, 2004). Alterations to the gut microbiome during active inflammatory bowel disease (Crohn's), with potential implications to the progression of the disease have been reported (Stange & Schroeder, 2019). Interactions between the microbiota and tapeworm infections are yet to be reported in the horse.

1.6.4 Current diagnosis

Diagnosis of tapeworm infection is difficult as infected horses do not show specific clinical signs unlike in horses with high levels of infection by strongyles resulting in weight loss and / or diarrhoea. Current methods for the diagnosis of tapeworm infection in the horse include coprological methods; faecal egg counts (FEC), floatation methods and sedimentation. Coprological methods lack sensitivity because of the irregular shedding of eggs by the parasite which are often in clumps that do not evenly distribute in the faeces. This results in a low recovery of eggs and an increase in the likelihood of false-negatives (Nielsen, 2016; Lyons *et al.*, 2018). The sensitivity of coprological techniques range from 2.8-75%, whereas the specificity ranges between 98-100% (Proudman & Edwards, 1992; Nilsson *et al.*, 1995; Tomczuk *et al.*, 2014, 2015). Eggs are more likely to be shed the higher the tapeworm burden (Nilsson *et al.*, 1995; Lyons *et al.*, 2018). Furthermore, the time of year (stage of the parasite lifecycle) affects the ability to detect eggs in the faeces. Eggs are most likely to be present in faeces in winter and early spring (temperate climates) (Tomczuk *et al.*, 2015). According to Nielsen *et al*., (2015) coprological techniques can give good diagnostic results but can vary in

performance as well as being time consuming to perform and require large amounts of faeces (30-50g compared to the normal 3-5g for strongyle FEC).

PCR assays using parasite genomic DNA from faeces have been explored as potential method to diagnose tapeworm (Drogemuller *et al.*, 2004). The PCR assays have been successful in being able to differentiate between the three different species of tapeworm when 50 or more worms are present (Bohórquez *et al.*, 2015). However, these assays may have the same limitations as FEC techniques in that a large amount of faeces are needed. As reviewed by Nielsen *et al.,*(2015), refinement of the DNA extraction method may achieve better results as in both studies extraction of parasitic DNA was from faeces. These techniques may not have become widely available as the sensitivity is similar to other current tests.

The use of serological antibodies to detect A. *perfoliata* infection was first explored by Hoglund *et al.*, (1995) and investigated total IgG against crude scolex antigen using an enzyme-linked immunosorbent assay (ELISA). A second blood serum ELISA test was developed in 1996 and is available commercially (Proudman & Trees 1996b). The commercially available serological ELISA is based on the elevated antibodies to the secretory immunoreactive proteins 12 and 13 kDa specific for *A. perfoliata*, which are part of the sub-class of IgG(T) (Proudman & Trees 1996a).

However, considerable variation of serum antibody levels in horses with similar burdens has been shown (Kjaer *et al.*, 2007). Antibody levels have been shown to decline 28 days post treatment but in others have continued to remain raised for several months (Proudman & Trees, 1996a; Abbott *et al.*, 2008). This may produce false positives when the antibodies are still raised but parasitic infection has cleared. The sensitivities and specificities reported for the commercially available serological ELISA test, based on post-mortem data are 68-70% and 71-95%, respectively (Proudman & Trees, 1996a; Kjaer *et al.*, 2007; Skotarek *et al.*, 2010). *Anoplocephala magna* is a less common equine tapeworm and cross-reactivity with this parasite play a role in the wide range of specificity values reported for the ELISA test. *Anoplocephala magna* resides in the small intestine, consequently it may have gone unnoticed in post-mortem studies (Bohórquez *et al.*, 2015). As described by Gasser *et al.*, (2005) the serological ELISA is 'a complementary tool to support clinical diagnosis, for research purposes and for the prevention of colic associated with *A. perfoliata*'. Applications of the serological test for research have demonstrated the following findings: levels of the anti-12/ 13 kDa IgG(T) antibody decreased as the risk of colic decreased post treatment with anti-cestode drugs in a horse population (Proudman & Holdstockt, 2000). Furthermore, age-

related correlations with antibody levels and infection intensities have also been described (Proudman *et al.*, 1997).

More recently a saliva ELISA test for tapeworm has been developed and marketed for routine diagnosis of *A. perfoliata* (Lightbody *et al.*, 2016). It has an advantage over the blood serum ELISA as it is a less invasive technique (but requires the horse to fast for 30 minutes prior to saliva collection) and does not require a veterinary surgeon to carry out the procedure. However, it is based on similar principles as the blood ELISA test and has similar sensitivity and specificity values of 85% and 83%, respectively. The antibody clearance from the saliva in horses after praziquantel administration was reported to be within 5-6 weeks. This is at the lower end of antibody clearance of that reported in the literature for the serological test (Proudman & Trees, 1996a; Abbott *et al.*, 2008). Lightbody *et al*., (2016) suggest that antibody clearance from the saliva is faster than from blood and therefore can be more effective at detecting current worm burden. However, this theory is only speculative, and the data is based on small sample size (11 horses) compared to more extensive studies performed on the serological ELISA (Proudman & Trees, 1996a; Abbott *et al.*, 2008). The use of the saliva ELISA test in a group of horses over 10 months lead to a reduction in anthelmintic usage of 86%, without an increase in tapeworm infection (according to the test) (Lightbody *et al.*, 2018). Further studies to test the efficacy of the saliva ELISA in larger cohorts of horses are required to fully determine whether limitations are the same as reported for the serological ELISA. Therefore, at present the current gold standard for diagnosing tapeworm infection is at post-mortem examination, which is not suitable for routine testing.

1.6.5 Treatment

Drugs targeted against nematode infections are ineffective against tapeworm (Nilsson *et al.*, 1995). There are two anthelmintic drugs available for the treatment of equine tapeworm infection; pyrantel pamoate and praziquantel. A dose rate of 6.6 mg base/g pyrantel pamoate was effective at removing 88% of *A. perfoliata* worms, confirmed at necropsy (Lyons *et al.*, 1989). Later work by Reinemeyer *et al.,* (2006) found that a higher dose rate of 13.2 mg base/kg pyrantel pamoate was effective at removing 95% of *A. perfoliata* worms. Administering pyrantel in low doses to horses daily in feed resulted in a disappearance in tapeworm, compared to the control which remained infected (Greiner & Lane, 1994). When determining the effect of praziquantel treatment, post slaughter Lyons *et al.,* (1995) found that 0.5 mg .kg^{-1} was effective against 0-100% of worms.

Current diagnostics are suitable for monitoring herd levels of tapeworm infection but not for determining real-time worm presence in individuals (Hoglund *et al.*, 1995). Therefore current advice to manage tapeworm infection in horses is to treat all horses on a premises once or twice per year following the grazing season e.g. autumn/ early winter (Nielsen *et al.*, 2013). However, prophylactically treating horses with unknown low burdens increases the proportion of the parasite population exposed to drugs and is likely to speed up the development of drug resistance.

1.6.6 Parasite resistance to anthelmintic drugs

Resistance in other groups of equine parasites to current drugs is an emerging problem in all species and is just as likely to occur in cestodes, especially given the prophylactic use of current anti-cestode drugs (Nielsen, 2016). Avoiding the treatment of animals with low burdens (below the threshold of causing disease), means greater numbers of the parasite population remain 'in refugia' and are able to dilute the population of resistant worms (Van Wyk, 2001). The monitoring of drug resistance in cestodes is not currently feasible in the live horse population as there is not a reliable diagnostic test available (Nielsen, 2016). Therefore, investigations into developing a new diagnostic test for detecting *A. perfoliata* are required.

1.7 Introduction to volatile organic compounds

Volatile organic compounds (VOCs) are a large group of carbon-containing molecules (with up to C20), which may be of biological or synthetic origin (Zhang & Li, 2010). The low molecular weight (50–200 Daltons) and high vapour pressure of VOCs allows them to enter the gaseous phase at room temperature, contributing to the odour of faeces, urine, breath, saliva, blood and sweat (Rowan, 2011). These compounds may be generated by physiological processes (including response to infection and inflammation) from the host or by its microbiota (Korpi *et al.*, 2009; Purkhart *et al.*, 2011). Furthermore, microbial produced VOCs may also be biochemically altered by the host. VOCs may also originate from components of the diet, from drugs or environmental contaminants (Pleil *et al.*, 2013; de Lacy Costello *et al.*, 2014). VOC analysis of faeces may provide a simple approach to understanding pathological changes (including inflammation) in the latter regions of the gut as demonstrated in mice and humans (Ahmed *et al.*, 2016; Reade *et al.*, 2019) gene. Figure 1.3 shows some of the VOCs that may be detected in the gaseous headspace of equine faeces that may arise from digestion, including from microbial fermentation (Amann *et al.*, 2014). To date, few studies have attempted to characterise the equine faecal VOC metabolome and influencing factors.

However initial studies have shown the groups of volatiles most predominant in equine faeces are acids, alcohols and ketones, likely to be derivatives of microbial fermentation (Proudman *et al.*, 2015). In terms of the faecal volatile metabolome only 5.49 and 8.54% of volatile organic compounds (VOCs) were shared between individuals pre- and post-supplementation with amylase-rich malt extract, respectively (Table 1.3) (Proudman *et al.*, 2015). This demonstrates that the equine faecal metabolome may be diverse in nature. Furthermore, dietary change can affect the pattern of faecal VOCs in the cow, horse and goat. (Zhao *et al.*, 2014; Fischer *et al.*, 2015; Proudman *et al.*, 2015; Snalune *et al.*, 2019). In the horse specifically, the introduction of spring pasture altered the faecal metabolome (Snalune *et al.*, 2019). It was speculated by the authors in the latter study that the decreased median abundance of methanol, isoprene and acetone in the faces after 6 weeks of spring pasture was a consequence of altered fermentation in the gut.

	Number of horses	**Total number of VOCs**	**Number of VOCs shared between all horses**	**% VOCs shared between all horses**
Pre-supplementation	8	61	9	5.49
Post-supplementation	6	61	14	8.54

Table 1.3 Numbers of volatile organic compounds shared in horse faeces pre- and post-supplementation with amylase-rich malt extract. Adapted from Proudman *et al*., (2014).

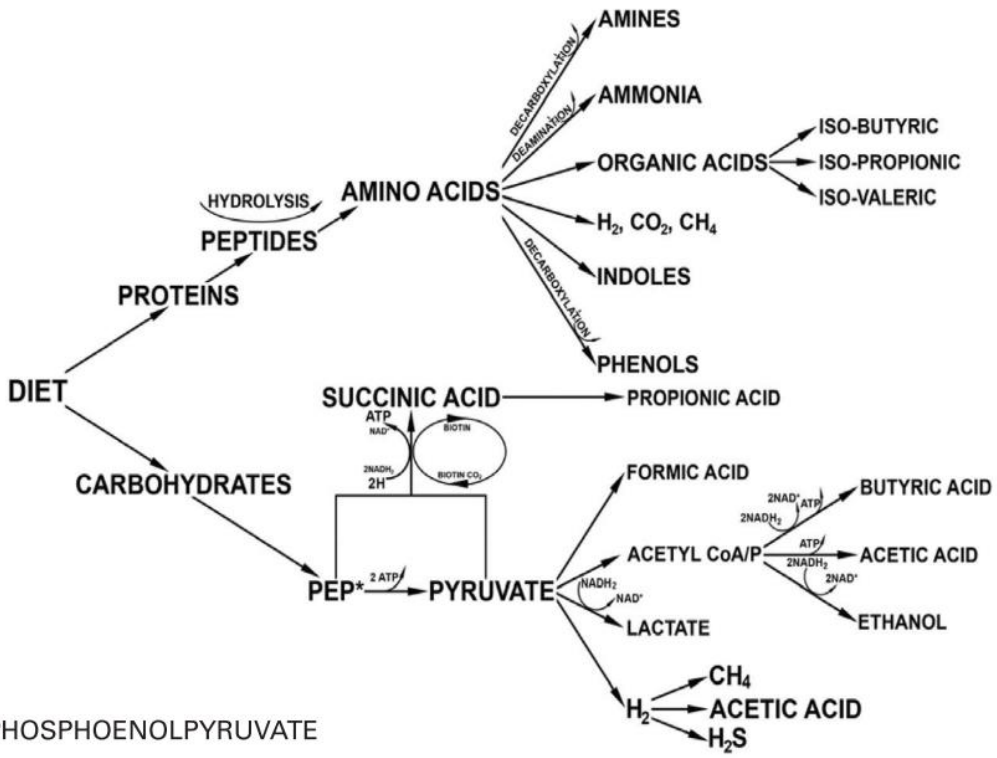


Figure 1.3 Source of volatile organic compounds from the equine diet and digestion. Taken from Cichorska *et al.*, (2014) (open access) who previously modified from Cummings & Macfarlane, (1991). Link to open access terms of use licence: https://creativecommons.org/licenses/by-nc-nd/3.0/legalcode

1.7.1 Volatile organic compounds as biomarkers for gastrointestinal disease

Alterations in faecal VOC profiles have been shown to be a non-invasive indicator of gastrointestinal disease in humans and other animal species (Garner *et al.*, 2007, 2008). An altered faecal VOC metabolome may be as a result of gut dysbiosis or changes in bacterial growth as a result of pathogens (Bergmann *et al.*, 2015). Horses with SCOD colic on one study were found to have greater faecal concentrations of acetone and methanol than non-colic controls (Turner *et al.*, 2013). Specific forms of colic that are thought to develop as a result of excess gas production and distention of the colon, such as SCOD have been associated with gut dysbiosis (Daly *et al.*, 2012). However, Turner and colleagues did not characterise the faecal microbiota to confirm whether alterations in bacterial communities also occurred.

Furthermore it is unsurprising that, as with the faecal microbiome, parasites can also alter the faecal VOC metabolome (Bond *et al.*, 2015). In pigs, oleic acid was found exclusively in individuals with *Trichuris suis* infection, along with altered fatty acid absorption (Li *et al.*, 2012). Li and colleagues also found through 16S rRNA gene-based and whole-genome

shotgun (WGS) sequencing that Proteobacteria and Deferribacteres changed in abundance in infected pigs. In mice infected with schistomes, 12 urinary and 5 faecal metabolites were identified as biomarkers and were able to differentiate between infected and non-infected mice (Li *et al.*, 2011a).

Investigations into the effect of the equine tapeworm, *A. perfoliata* on the equine hindgut microbiome and metabolome have not previously been carried out. Identification of a faecal biomarker for *A. perfoliata* has the potential to be a valuable and much needed diagnostic test for the parasite that does not rely on current diagnostic tests that are reliant on host immunological changes which are slow to respond to changes in parasite burdens. The use of metabolomics may provide a non-invasive, cost-effective and improved diagnostic test for monitoring parasite infections, enabling accurate identification of horses that require anthelmintic treatment. This would reduce the risk of parasite-associated disease and would reduce unnecessary use of anthelmintics in horses with low worm burdens, thereby minimising the risk of anthelmintic resistance developing. In addition, together with microbiome analysis, metabolome studies would increase our understanding of the biological mechanisms underpinning development of parasite-associated colic. Two recent reviews highlighted the importance of studying the metabolome in addition to the microbiome for parasite-microbiota interaction to expand knowledge of microbial function (Peachey *et al.*, 2017; Rapin & Harris, 2018).

1.8 Hypothesis, aims and objectives

Hypothesis: Knowledge of the longitudinal microbiome and VOC profiles of different cohorts of healthy horses will generate baseline data that can be used to assist diagnosis of horses at high risk of different types of gut-related disease. The VOC profile and microbiome of horses infected with *A. perfoliata* will differ from that of healthy horses. These studies will enable a greater understanding of the VOC metabolome and microbiome in different populations of horses and provide evidence that VOCs can act as a proxy to the microbiome.

The aim of this PhD thesis is to carry out a series of investigations into the faecal and hindgut microbiome and metabolome of horses. The studies were carried out in the form of temporal sampling in two cohorts of healthy horses and in a third cohort where horses were identified as either positive or negative for *A. perfoliata* post-mortem.

Objectives:

- To establish an optimal method for extracting VOCs from equine faeces using HS-SPME-GCMS.
- To characterise the temporal faecal VOC metabolome in populations of healthy horses, specifically around the time of foaling and in horses maintained at pasture all year round.
- To characterise the faecal mycobiome of grazing horses at four time points representing seasonal change.
- To compare the colonic microbiome (16S rRNA sequencing) and VOC metabolome (HS-SPME-GCMS) of horses infected with *A. perfoliata* and non-infected controls.
- To identify a faecal VOC biomarker of *A. perfoliata* infection in horses.
- To integrate VOC metabolome and sequencing data where possible to strengthen evidence of VOCs as a proxy for the bacterial microbiome and mycobiome.

Chapter 2 Methodology: general methods and method optimisation

2.1 Introduction

The metabolome is described as the total number of small molecules (metabolites) that are found within an organism, cell or tissue (Fiehn, 2002). As discussed in the previous chapter, volatile organic compounds (VOCs) are a component of the metabolome of the intestinal bacteria. The study of the equine hindgut bacterial populations is an active area of research, as it is regarded to be extremely important to equine health (Daly *et al.*, 2012). VOC analysis of faeces provides a simple approach to understanding functional changes of the microbiota of the distal intestine. Furthermore, VOCs offer a cost-effective alternative to microbiome studies. The faecal VOC profiles of horses have been found to alter in response to dietary supplementation, from the ingestion of probiotics, and when suffering from gastrointestinal disease (colic) (Turner *et al.*, 2013; Ishizaka *et al.*, 2014; Proudman *et al.*, 2015).

The bacterial metabolites of the equine hindgut have received much less attention than the bacteria themselves. However, studying the metabolome and microbiome in synchrony will give a wider picture of how the bacteria are functioning and the role they play in their contribution to nutrition and health of the animal. To date few studies of the equine hindgut microbiota have combined VOC and bacterial sequencing of the same samples (Proudman *et al.*, 2015). As mentioned in a recent review, work combining characterisation of the intestinal microbiota and the metabolites produced is needed for greater understanding of underlying mechanisms related to health and disease in mammals (Brosschot & Reynolds, 2018).

The first aim of this chapter is to outline the methods that may be applied to characterise the metabolome and microbiome and describes the resultant methods chosen for this thesis. The second aim is to develop optimal preparation steps, specific to horse faeces, for the metabolomics platform to be applied in this thesis.

2.2 Methods to sample the equine hindgut

Several methods can be employed to sample the hindgut contents. A summary of options and a justification of the methods chosen for this thesis are as follows:

2.2.1 Surgical procedure

Ethical restrictions mean surgical procedure can only be carried out on animals undergoing abdominal surgery as standard veterinary practice and is therefore not suitable for characterising the microbiome and metabolome of healthy horses. Repeat sampling is not possible unless the animal needs repeated procedures. Sampling the hindgut in this way would restrict the number of animals that could be recruited for study.

2.2.2 Cannulated animals

Cannulation involves surgically inserting an access portal to allow repeated sampling of the hindgut contents in live animals. However, cannulated animals are expensive to maintain which results in very small sample sizes. Cannulation can have a negative effect on horse-health including: fluctuation in weight, abrasions and leakage around the cannula site and hair-loss (Venable *et al.*, 2017a). It is also not known whether the impact of cannulation itself influences the gut microbiota. Furthermore, keeping cannulated animals for research is not permitted in the UK.

2.2.3 Direct removal of faeces from rectum

This method avoids contamination, that may occur from the environment, of a naturally-passed sample, e.g. from the pasture/soil or stable bedding material, and limits sample exposure to an aerobic environment. This method is invasive and can only be performed under standard veterinary procedure. There is also a risk of rectal tears.

2.2.4 Naturally-voided faeces

This is non-invasive, does not require a Home Office project licence and allows many samples to be collected from the same animal. The samples can be collected easily without training or specialist equipment. As demonstrated by 16S rRNA targeted terminal restriction fragment length polymorphism (TRFLP), 454 pyrosequencing and MiSeq sequencing, faeces reflect the microbial populations of the distal regions of the gastrointestinal tract. The right dorsal and left dorsal colons were represented by the faeces more so than the caecum (Dougal *et al.*, 2011a, 2011b; Costa *et al.*, 2015a).

2.2.5 Intestinal contents post mortem

Horses cannot be killed for experimental purposes. It may be difficult to obtain information on diet or medical history for samples sourced from abattoir material. Sampling cannot be repeated but it does allow a gold standard diagnosis of tapeworm infection and enables gut contents to be sampled by directly evaluating parasite burdens in the intestinal lumen.

2.2.6 Chosen sampling methods

For much of the work in this thesis, naturally-voided faeces were the chosen sampling medium. Samples were collected from healthy horses over time, as repeated sampling from the same animals was required. An investigation into the effect of the parasite *Anoplocephala perfoliata* was carried out as detailed in later chapters. For this, hindgut (pelvic flexure) and rectal contents were sourced from abattoir material to enable accurate determination of burden. Luminal gut contents were considered the most accurate medium for determining the microbiome of this area. For biomarker investigations, rectal contents from these animals were collected because samples for biomarker studies need to be non-invasive and easily collected from live horses to be of use.

2.3 Methods to extract volatile organic compounds from faeces

This section will explore the methods that can be used to extract VOCs from biological samples. Traditionally, solvents were used for VOC extraction but these require long extraction times and many preparation steps, not to mention the environmental impact of solvent use (Zhang & Li, 2010). Fortunately, many solvent-free techniques are now available and some of the main gas phase extraction and sorbent extraction techniques will be discussed here. An exhaustive review of extraction methods is beyond the scope of this thesis and further reviews can be found elsewhere (Koning *et al.*, 2009; Jochmann *et al.*, 2014).

Solid phase micro-extraction (SPME) involves a silica coated rod (fibre) entering either directly into a sample (immersion) or into the headspace of a sample. The headspace is referred to here as the region of air above a sample. Both methods share the common aim of quantifying target analytes (Pawliszyn, 2000). The lifetime of a fibre is extended and background noise is reduced (as components of the faecal matrix will not be extracted) when inserted into the headspace of a sample rather than direct immersion. Immersion of the fibre can only be carried out in aqueous samples free from carbohydrates, protein and fat (Roberts *et al.*, 2000). When exposed to the headspace, analytes are transported through the fibre

coating. The coating material and thickness has a direct impact on the analytes adsorbed or absorbed and so it is a crucial factor to consider in experimental design. SPME is not an exhaustive technique (it cannot extract the entire amount of a compound being emitted from a sample), but the extraction is considered complete when equilibrium is reached between the fibre and the sample. Once equilibrium is reached, continuing to expose the fibre will not obtain more analytes. For complete equilibrium to happen there are three main phases: i) fibre coating to sample, ii) headspace to sample and iii) fibre coating to headspace. Equilibrium can be represented by the following equation:

$$n = \frac{K_{fs}\, V_f V_s C_0}{K_{fs} V_f + V_s}$$

n = the mass of analyte extracted by the coating

K_{fs} = fibre coating–sample matrix distribution constant

V_f = fibre coating volume

V_s = sample volume

C_0 = the initial concentration of a given analyte in the sample

Purge and trap (P&T) involves a sample being enclosed within a vessel which is purged with an inert gas. Following this, the VOCs are pushed into an adsorbent trap and the inert gas escapes through a vent. The trap is then heated to desorb the compounds which enter the gas chromatograph (GC) via a carrier gas. There are several types of sorbent commercially available including Tenax®TA and Carbotrap that can be used. In addition to the sorbent, trap and purge time and temperature can be altered to achieve high selectivity for different biological VOCs (Koning *et al.*, 2009; Woolfenden, 2010). P&T has been successfully used to extract VOCs from human faeces but has not previously been applied to equine faeces (De Preter *et al.*, 2009).

Thermal desorption (TD) tubes are like P&T in that a sample is enclosed within a vessel. The vessel may be made of materials including glass, stainless steel and coated steel (Woolfenden, 2012). The procedure differs from P&T in that the sample is heated in a flow of inert carrier gas which volatiles desorb into following injection into a GC. Thermal desorption can be a single (as described above) or a two-stage process. The two-stage process involves a second narrower tube which contains a focusing trap (sorbent) which is heated and concentrates the analytes further, resulting in sharper GC peaks (Woolfenden,

2012). As with P&T, the choice of sorbent is very important as some are better at trapping low or high molecular weight compounds (Turner, 2016). Purchase of the TD tubes can be initially expensive; however they can be reused hundreds of times (Woolfenden, 2012). The use of TD tubes to extract equine faecal VOCs has been successfully applied (Proudman *et al.*, 2015).

P&T and TD tubes have advantages over SPME in that they can be fully. However, drawbacks of P&T and TD tubes are that both require an expensive analytical instrument and automated systems for high sample throughput, and so they are less common than alternative methods e.g. SPME (Jochmann *et al.*, 2014). There are pros and cons for each of the extraction techniques discussed here (summarised in Table 2.1): it is important to select the method most suitable for the application intended.

Extraction technique	**Advantages**	**Disadvantages**	**References**
SPME	Common technique applied to faeces, sensitive, few preparation steps and comparatively low cost.	Difficult to quantify, competition for binding between analytes risk of low analyte recovery or long extraction times and the fibres are fragile.	(Povolo & Contarini, 2003; Karu *et al.*, 2018)
TD tubes	Sensitive, samples can be stored in tubes for prolonged periods.	Sorbent selection important, quantification difficult.	(Woolfenden, 2010)
P&T	Exhaustive extraction possible, quantifiable, very sensitive.	Water vapour can interfere with analysis, careful optimisation of purge time and temperatures needed.	(Koning *et al.*, 2009)

Table 2.1 Advantages and disadvantages of solvent-free extraction techniques. Adapted from (Turner, 2016)

2.4 Methods to characterise the hindgut metabolome

Suitable techniques to characterise the hindgut and faecal metabolome can be targeted or untargeted. A targeted approach requires metabolites to be pre-determined, it can be highly quantitative and data processing is much less laborious than untargeted approaches (Roberts *et al.*, 2012). In terms of equine studies, there have been few that have used an untargeted metabolomics approach (Turner *et al.*, 2013; Escalona *et al.*, 2014; Proudman *et al.*, 2015). Many studies, for example work by Biddle *et al.*, (2013); Blackmore *et al.*, (2013); Hansen *et al.*, (2015) have quantified specific volatile fatty acids (VFAs) using various techniques. It is useful to study fluctuations in VFAs to understand the effect of a range of factors including dietary change and disease status. However, the benefits of an untargeted study are that a much bigger picture of the entire metabolome is obtained, not just the main VFAs but also a range of metabolites that are products of microbial digestion. An example is indole which is a product of protein degradation by proteolytic bacteria. An untargeted approach is also the first step in biomarker discovery investigations, which is a main aim for this thesis. Therefore, it is established that an untargeted approach should be taken and the suitability of untargeted methods for characterising the equine faecal metabolome will be discussed here. The main techniques for untargeted metabolome analysis will be discussed briefly; more detailed reviews can be found elsewhere (Primec *et al.*, 2017; Cameron & Takáts, 2018).

Gas chromatography (GC) coupled with a mass spectrometer (MS) involves a sample being injected into a GC oven, in which there is a column (stationary phase). The injected sample is carried through the column via a carrier gas (mobile phase) of helium, nitrogen or hydrogen. Compounds elute at a retention time (RT) which is determined by several factors including polarity, boiling point, and column material. The eluted compounds then enter the MS which separates metabolites based on their mass to charge ratios (*m/z*). The molecules are bombarded by electrons from an electron ionization (EI) source and are fragmented into ions. The fragmented metabolites are then selected based on their m/z ratios and pass through a detector which allows the abundance of the ions to be determined (Murayama *et al.*, 2009).

An alternative separation technique to GC, often coupled to MS, is liquid chromatography (LC). As the names suggest, LC uses liquid as the mobile phase whereas GC uses a carrier gas. Gas chromatography systems have a higher operating temperature than LC which is one of the reasons (along with column properties) why LC can separate compounds with higher molecular weights than GC (Snyder & Kirkland, 1979). The chromatography technique selected (GC or LC) depends on the compounds of interest, i.e. low molecular weight or high

molecular weight, or the techniques can be used to complement each other to detect a much broader range of metabolites.

Nuclear magnetic resonance spectroscopy (NMR) is based on the magnetic spin properties of certain nuclei, such as ^{1}H, ^{13}C, ^{31}P and others. Strong magnetic field and radio frequency (RF) pulses are applied to the nuclei. This presence of a magnetic field causes the nuclei to spin and the RF energy absorbed causes a change from a low to high energy spin state. The radiation emitted during the relaxation process is detected and creates an NMR signal (Dunn & Ellis, 2005). NMR may be one, two or three dimensional; the theory and the extensive applications of NMR are reviewed in-depth elsewhere (Kruk *et al.*, 2016). NMR is highly selective and can provide complex information on the structure of molecules, making accurate identification a reproducible technique.

2.5 Selected methods to characterise faecal the metabolome: Headspace solid phase micro-extraction gas chromatography mass spectrometry (HS-SPME-GCMS)

There are multiple options for the extraction of VOCs from faeces. The method chosen here is headspace SPME. The latter technique is simple, involving very little sample preparation, is portable, if needed, and is cost effective. The number of samples that can be processed in a 24-hour window can be increased by using an automated system which does not require manual injection by an operator. There is no extraction technique that will detect all biological VOCs, therefore multiple techniques should be selected to identify as many as possible. Previously TD has been applied to horse faeces when coupled to GCMS, whereas the increasingly popular SPME has received little application to horse faeces and therefore has been chosen for this work. Furthermore, different SPME fibre coatings (Table 2.2) targeting a wide range of analytes are available commercially. Although compound quantification using SPME can be difficult, it is suitable for compound discovery studies such as those to be carried out in this thesis.

GCMS remains one of the most widely used platforms for identifying VFAs and has been selected as the chosen method for this work to characterise the equine faecal metabolome (Primec *et al.*, 2017). The decision has been made based on the high sensitivity of the technique, the comprehensive mass spectral databases that are available and the ability to analyse a high number of samples at a more economical rate compared to similar platforms. GCMS appears to extract a greater number of metabolites when applied to equine faeces; 81

volatiles were identified compared to 43 and 16 when NMR and SIFT-MS were used, respectively (Turner *et al.*, 2013; Escalona *et al.*, 2014; Proudman *et al.*, 2015).

The use of GCMS provides detailed metabolomic profiling that is required for this work, including identification of biomarkers of tapeworm infection, compounds may be host response, parasite specific or from gut bacteria.

Fibre coating	Film thickness (µm)	Polarity	Coating stability	Maximum temperature (°C)	Analytical application	Recommended uses
Polyacrylate (PA)	85	Polar	Cross-linked	320	GC/HPLC	Polar semi-volatiles (phenols)
Polydimethyl-siloxane (PDMS)	100 30 7	Non-polar Non-polar Non-polar	Non-bonded Non-bonded Bonded	280 280 340	GC/HPLC GC/HPLC GC/HPLC	Volatiles Non-polar semi-volatiles Mid to non-polar semi-volatiles
Carbowax-poly (ethyl glycol) (PEG)	60	Polar		250	GC/HPCL	Polar analytes, alcohols
Carboxen (CAR)-PDMS	75 85	Bi-polar Bi-polar	Cross-linked Cross-linked	320 320	GC GC	Gases and volatiles
Divinylbenzene (DVB)-PDMS (StableFlex fibre)	65 60 65	Bi-polar Bi-polar Bi-polar	Cross-linked Cross-linked Cross-linked	270 270 270	GC GC/HPLC GC	Polar volatiles General purpose Polar volatiles
CAR-DVB-PDMS	50/30	Bi-polar	Cross-linked	270	GC	Odours and flavours

Table 2.2 Commercially available SPME fibre coatings. Adapted from Mills & Walker, 2000; Risticevic *et al.*, 2010; Shirey, 2012.

2.6 Methods to characterise the hindgut microbiota

Culturing methods have been used to study bacterial populations of the equine hindgut (Respondek *et al.*, 2008). The findings of these studies are discussed in **Chapter 1**. However culturing methods are limiting, largely because of the anaerobic nature of the gut bacteria and our knowledge of the specific carbon sources and conditions bacteria require to grow. This has led to the development of alternative culture-independent molecular techniques and next-generation sequencing (NGS) technologies to study microbial communities. The main molecular and NGS technologies will be discussed in brief; there are more in-depth reviews available elsewhere (Zoetendal *et al.*, 2004; Oulas *et al.*, 2015; Besser *et al.*, 2018).

The 16S rRNA gene is made up of around 1500 nucleotides and is a gene shared by all bacteria. There are 9 hypervariable regions which can be targeted and amplified using PCR primers and used as markers to investigate bacterial community diversity (Head *et al.*, 1998). Regions V1- V6 have been used for studying the equine hindgut and faecal microbiota (Costa *et al.*, 2012; Dougal *et al.*, 2014). Regions V7-V9 are not recommended as identifications can become less reliable (Cai *et al.*, 2003).

Prior to the development of NGS technologies, molecular 'finger printing' techniques including denaturing gradient gel electrophoresis (DGGE) and T-RFLP have been applied to study differences in microbial community composition in the equine gut and faeces (Earing, 2012; Blackmore *et al.*, 2013). Furthermore other techniques including fluorescence *in situ* hybridization (FISH) and quantitative polymerase chain reaction (qPCR) have been applied to quantify known organisms in equine faeces and gut contents (Milinovich *et al.*, 2007; Shepherd *et al.*, 2014). Some of the main advantages and limitations of these techniques are summarised in Table 2.3.

Molecular techniques are still relevant and are valuable tools for studying the hindgut microbiota. In a direct comparison between T-RFLP and Ion Torrent personal genome machine (PGM), similar results were obtained for diversity and richness as well as sample clustering (de la Fuente *et al.*, 2014). The most obvious drawback with molecular techniques is that unknown organisms cannot be identified, whereas NGS is able to do so. Moreover, the arrival of NGS has revolutionised the ability to study microbial populations with far greater depth and precision utilising much less labour intensive and time-consuming protocols. The most common platforms include Ion Torrent PGM, Illumina and 454 pyrosequencing. The main principles of NGS technologies are to perform sequencing of millions of small fragments of DNA. Identification is then made by entering sequences into available databases, including the Ribosomal Database Project (http://rdp.cme.msu.edu).

For bacterial sequencing, the targeting of the variable regions of the 16S rRNA has been applied to equine faecal and hindgut contents since the emergence of NGS studies in this field around 2012 (Shepherd *et al.*, 2012; Steelman *et al.*, 2012). Next-generation sequencing has enabled a far greater understanding of the equine hindgut microbiota and the findings of these studies are discussed in **Chapter 1**.

Many of these molecular and targeted NGS techniques require DNA extraction and PCR amplification and therefore share similar limitations. These include variations between DNA extraction protocols and contamination during extraction (Shepherd *et al.*, 2015) and PCR

biases which may include amplification errors and preferential amplification (Meisel & Grice, 2016).

Whole-genome metagenomic shotgun sequencing (WGS) is a technology that can sequence all DNA present in a sample which is advantageous in that it does not require PCR amplification. In addition, WGS allows the re-construction of entire bacterial genomes which means previously unknown species can be added to reference databases (Meisel & Grice, 2016). Additional knowledge of entire bacterial genomes can also provide further information to understand the function of the organism (Meisel & Grice, 2016). If previously unidentified species are found during targeted 16S rRNA sequencing they are referred to as unclassified as targeted sequencing does not provide enough information for identification. The drawbacks of WGS are that it is more expensive and laborious for many samples than targeting the 16S rRNA gene. Whole-genome sequencing is very useful for more in-depth, species-level identification which may be interesting to pursue from trends that may arise from initial 16S rRNA experiments. To date, limited studies using WGS have been performed on equine hindgut microbiota (Kristoffersen *et al.*, 2016). Following initial 16S rRNA studies into dietary induced changes to the hindgut microbiota by Kristoffersen *et al*., a selection of samples was then subject to WGS. From this the authors were able to identify genes that were associated with the bacteria of interest which revealed further indication of their functionality including their fermentative characteristics. The findings of this work are discussed in more detail in **Chapter 1**. WGS is an active tool being applied to the gut bacteria of animals and humans (Oniciuc *et al.*, 2018). In the rumen the use of WGS has identified strains of *E. coli* with antimicrobial resistant genes (Rehman *et al.*, 2017). Further application of WGS to equine hindgut bacteria is likely to increase knowledge of bacterial function.

Despite the clear advantages of WGS, the technique is still restricted by DNA extraction bias. Furthermore DNA is representative of all organisms whether they were active in the gut or lying dormant (Schirmer *et al.*, 2018). Application of technologies in the study of gut bacteria include RNA sequencing (metatranscriptomics) which are more indicative of active species (Amrane *et al.*, 2018).

Targeted 16S rRNA studies have greatly advanced the field of study of the equine hindgut microbiota and there are still many applications for this technique in the field. However whole-genome and RNA sequencing technologies would provide more in-depth indications of species and organism function, knowledge of which is currently lacking for equine hindgut bacteria. The technique selection for characterising the hindgut bacteria is dependent on the

question being asked, because as discussed there are advantages and disadvantages with each tool.

Technique	Advantages	Limitations	References
FISH	Sensitive and quantitative, microbial in-situ identification	Only known species can be identified, enrichment steps often required	(Zoetendal *et al.*, 2004; Girones *et al.*, 2010)
T-RFLP	Low cost	Cannot identify species, not quantitative, low biodiversity resolution	(Clement *et al.*, 1998; Orcutt *et al.*, 2009; de la Fuente *et al.*, 2014)
qPCR	Sensitive and quantitative	Can only target and quantify specific species of organisms e.g. total anaerobic fungi.	(Edwards *et al.*, 2008)
DGGE	Allows extraction of band for amplification and sequencing, robust first line test.	Different bacterial species can display same separation. Time consuming	(Muyzer *et al.*, 2004; Tardy *et al.*, 2009; El Sheikha *et al.*, 2012)
Phylogenic microarray	High biodiversity, quantitative	Cannot detect novel species, expensive	(Paliy & Agans, 2012)

Table 2.3 Comparison of molecular techniques for studying bacterial populations.
Adapted from (Maccaferri *et al.*, 2011)

2.7 Selected methods to characterise the microbiome: targeted 16S rRNA MiSeq next generation sequencing

Targeted sequencing of the 16S rRNA gene has been chosen for this thesis because of the research question being asked. A comparison of the hindgut bacterial communities of tapeworm and non-tapeworm infected horses has not previously been undertaken. In fact, few studies have addressed the question of the impact of intestinal parasites on the equine hindgut bacteria (Clark *et al.*, 2018; Peachey *et al.*, 2018, 2019). Therefore, the targeted 16S rRNA approach is deemed more suitable than WGS which would be expensive and time-consuming for a first exploration study. Molecular finger printing techniques would provide diversity and richness but would not be able to identify the organisms present. One of the main aims of this thesis is to identify bacteria that may be associated with tapeworm infection and their correlated metabolites.

The MiSeq sequencing platform is a suitable choice for this work. It provides an output of similar quality to 454 pyrosequencing but at a much lower cost (Fadrosh *et al.*, 2014). A criticism of MiSeq technology is that it does not sequence low sequence diversity samples well e.g. those with few species of bacteria. However previous equine microbiota profiling

has revealed a bacterial population of high diversity but low abundance (Dougal *et al.*, 2013). MiSeq technology has been successfully used to sequence bacterial DNA from equine caecal contents (Moreau *et al.*, 2014) and faeces (Leng *et al.*, 2014) and therefore is a justified sequencing platform for the current work.

2.8 General metabolomics methods used in this thesis

Methods to study the metabolome are applied to all chapters in this thesis. Methods to study the microbiome are specific to **Chapters 3** and **5** and are detailed in those chapters.

2.8.1 Headspace-solid phase micro extraction (HS-SPME)

VOCs were extracted from the headspace of faeces when the SPME fibre was exposed to the headspace of a 10ml or 20ml glass vial as programmed using a CombiPal auto-sampler (CTC Analytics, Switzerland). The extraction temperature was 60^{o}C for 20 minutes. Prior to extraction, vials were incubated at 60^{o}C for 30 minutes. The SPME fibres used were divinylbenzene-carboxen-polydimethylsiloxane (DVB-CAR-PDMS) 50/30µm (1cm) or carboxen-polydimethylsiloxane CAR-PDMS (85µm) (Sigma-Aldridge, Dorset, UK). The fibres were pre-conditioned in accordance with manufacturer's instructions before use.

2.8.2 Perkin Elmer Clarus 500 GCMS analysis conditions

GCMS analysis was carried out using a Perkin Elmer Clarus 500 Quadruple bench top system (Beaconsfield, UK). Separation of VOCs was performed in a Zebron ZB-624 GC column with an inner diameter of 0.25mm, length of 60m and a film thickness of 1.4µm (Phenomenex, Macclesfield, UK). Helium of 99.996% purity was used as a carrier gas, set at a flow rate of 1ml/min (BOC, Sheffield, UK). SPME fibre desorption temperature and time were 220^{o}C and 5 minutes, respectively. The GC oven was initially set at 40^{o}C and held for 1 minute before rising to 220^{o}C at a rate of 5^{o}C per minute and held at 4 minutes (total run time of 41 minutes). The MS was operated in electron impact ionization EL+ mode, scanning ion mass fragments from 10 to 300 *m/z* with an inter-scan delay of 0.1 seconds and a resolution of 1000 FWHM (Full Width at Half Maximum).

Samples were analysed in triplicate to keep random instrument or handling errors to a minimum and to partially correct for false positive and false negative findings (Bader *et al.*, 2016). The order of running samples was computer randomised to reduce run-time bias (Dunn *et al.*, 2012). Laboratory air was sampled frequently to rule this out as a possible

contaminant source. Blank vials were tested between samples to ensure VOCs were originating from faeces and to prevent carry-over of which may originate from the fibre or the column (Hughes *et al.*, 2007).

2.8.3 Reference solution

To ensure consistent running of samples when several weeks of analysis were required, a reference solution was analysed once per week. A total of 10 µl of 2-pentanone, benzaldehyde and pyridine and 10 mg of indole were added to 500 ml of distilled water and dissolved. This was then stored in the fridge (4^{o}C) with a shelf life of three weeks. Aliquots of 100 µl were added to 10 ml vials and were analysed using HS-SPME-GCMS and run in triplicate. The method was as detailed in sections 2.8.1 and 2.8.2. Reference solution vials were not heated prior to, or during SPME.

2.8.4 Metabolomics data analysis

2.8.4.1 Data processing and VOC identification

Raw files were converted into the required net CDF form using DataBridge software (PerkinElmer), prior to data processing. The data were processed using Automated Mass Spectral Deconvolution System (AMDIS-version 2.71, 2012). The original factory settings for AMDIS were used with the following exceptions: a minimum match factor of 70 and the analysis type was simple. Deconvolution settings had a component width of 12 and adjacent peak subtraction was set to one. Resolution, sensitivity and shape requirements were set to medium, high, medium, respectively. The National Institute of Standards and Technology (NIST) mass spectral library (version 2.0, 2011) was used to putatively identify VOCs which were added to an in-house built library. The selection of VOCs added to the library was based on ion fragment patterns, R match and probability percentage (>70%). Random checks of processed data were compared to original chromatograms to check compound matching was consistent and precise. Three compounds included in the reference solution mentioned in 2.8.3 (2-pentanone, benzaldehyde and indole) were also part of the equine faecal VOC metabolome and we were able to confirm the findings of these compounds in the NIST library. This work is consistent with the (MSI) guidelines for identifying compounds (Sumner *et al.*, 2007).

The R package Metab (Aggio *et al.*, 2011) was used to align data and, at this stage, an average of triplicate-run samples was taken. Pooling technical replicates by taking an average is common practice with metabolomics data. However, there is a lack of published standard

guidelines with how to deal with missing data when taking an average. For this thesis a method was adapted based on methods used by published studies for the removal compounds present in very few samples and imputation of half-minimum values (Khalid *et al.*, 2015; Wei *et al.*, 2018). The method for taking averages of technical replicates in this thesis was defined by the following criteria:

- When a compound was present in all three samples, an average was taken.
- If a compound was present in two out of three samples, a half-minimum value was imputed for the third sample and an average taken.
- If a compound was present in just one out of three triplicates, it was removed.

Once an average of triplicates was attained the removal of low-quality data points (compounds present in few samples) was carried out on the data matrix. Compounds present in very few samples may be individual specific or have a very low detection level. These compounds may also be a result of random instrument errors or missed during peak-picking (Bijlsma *et al.*, 2006). The removal of low quality data points is important for reducing downstream false discovery rates (FDR) (Xia *et al.*, 2012). For a compound to remain in the analysis it had to be present in a minimum of 50% of samples within at least one condition. Following low quality data point removal, missing data still present were replaced with a half minimum value of that condition.

Following false positive filtering and missing data imputation, the data were normalised using log transform (natural log) to enable even distribution of the data. This makes trends more easily identifiable and enables parametric statistical analysis to be carried out. An example of the spread of data pre and post normalisation is shown in Figure 2.1.

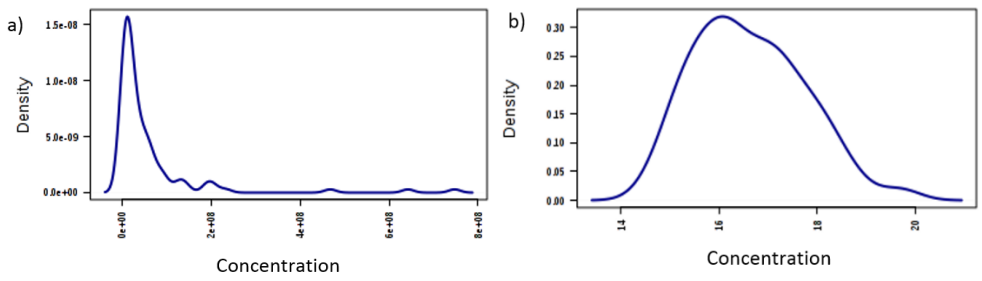


Figure 2.1 The spread of an example set of data used in this thesis (data in section 2.9.3.2) pre a) and post b) log transformation.

2.8.4.2 Metabolome statistical analysis

The software package R, version 3.1.2, IBM SPSS statistics, Microsoft Excel® and the online software tool Metaboanalyst (Xia *et al.*, 2012) were used to perform statistical analysis. Benjamini-Hochberg to control false discovery rate (FDR) was applied to adjust p-values for multiple comparisons.

i) Univariate methods

Univariate analysis was performed to identify changes in the number of VOCs and absence/presence (Fishers' exact test) of VOCs between groups. For these questions the analysis was carried out on data prior to imputation of missing values. The number of VOCs detected within samples was tested for normal distribution using the Shapiro-Wilk test, which determined the appropriate statistical test to use. For normally distributed data a t-test or a one-way analysis of variance (ANOVA) was applied. For data not normally distributed a Kruskal-Wallis test followed by pairwise comparisons were used. The coefficient of variation (CoV) of VOC peak area was applied where appropriate to assess the reproducibility of the data.

For the mean abundance of compounds (peak area) univariate analysis was performed to identify trends in individual compounds that may be up or down regulated between groups. The statistical tests used were either a t-test or ANOVA followed by Tukey's HSD test for pairwise comparisons.

ii) Box and whisker plots

Box and whisker plots (boxplots) were constructed to visually represent the spread of data (range). The length of the box is made up of the first quartile (Q1) to the third quartile (Q3). Between Q1 and Q3 is a line known as Q2 which represents the median. The distance between Q1 and Q3 is defined as the interquartile range (IQR). Whiskers extend from either Q1 or Q3 to represent the smallest and largest non-outliers, respectively. Specifically, outliers were identified as either larger than Q3 by at least 1.5 times the IQR, or smaller than Q1 by at least 1.5 times the IQR.

iii) Principal component analysis (PCA)

Principal component analysis was conducted on samples to provide a visual representation of the variation between its principal components (the underlying structure of the data). Two variables are combined into one to allow ease of comparison. PC1 accounts for the most variation in the dataset and is constructed from the major eigenvector when two variables

are plotted against each other. PC2 is the minor eigenvector and represents much less variation as the points are plotted much further away from it. Principal component analysis reduces the dimensionality of data which allows the scatter of it to be viewed easily. The PCA plots were constructed in R using the package ggplot2. Dr Shebl Salem (University of Liverpool) wrote the PCA script.

iv) Sparse partial least squares discriminant analysis (PLS-DA)

Sparse partial least squares discriminant analysis is a supervised classification tool. The algorithm differs from PCA in that the groupings are known before the data is projected into the model. In the supervised classification model used in this thesis the aim was to reduce the dimensionality of the data by performing feature selection (Lê Cao *et al.*, 2011). Supervised classification models can lead to over fitting. Overfitting is likely to occur for small sample sizes, if there are too many variables or if there is no actual difference between the groups. To check the integrity of the model, a method known as the balanced error rate (BER) was performed. The BER was calculated by the average proportion of wrongly classified samples weighted by the number of samples in each class (Rohart *et al.*, 2017). The lower the BER the better the performance of the model. Despite the possibility of overfitting sPLS-DA is useful because it is able to classify noisy data, therefore when used correctly it can be a very informative tool (Gromski *et al.*, 2015). sPLS-DA was performed in using the mixOmics R package (Rohart *et al.*, 2017). Dr Umer Zeeshan Ijaz (University of Glasgow) wrote the sPLS-DA scripts.

v) *Permutational analysis of variance (PERMANOVA)*

Permutational analysis of variance (PERMANOVA) following calculation of a Bray – Curtis dissimilarity matrix was used to estimate how much variation in the data could be explained by factors (e.g. individual horse, feed type and parasite burden). The R package vegan and function Adonis were used to carry out the analysis and Dr Shebl Salem (University of Liverpool) wrote the script. P values were not adjusted, as correction for multiple comparisons were included in the PERMANOVA model.

vi) Modelling VOC abundance for association with factors of interest

To investigate the association of VOC abundance with factors of interest (e.g. time, feed type, parasite burden) two types of statistical model were used. For longitudinal data, a linear mixed effect model (LME) was applied using the R package lme4. Individual horse was set as the dependent variable and other factors of interest as the independent variables. An LME model approach rather than rANOVA was used because LME can cope with missing data. For

comparisons between disease and control with multiple factors of interest, a multivariate multiple regression model was performed in R to evaluate the association of VOC abundance with factors of interest. Dr David Hughes wrote the LME script.

vii) Receiver operator characteristic (ROC) curves

One of the most objective and statistically valid methods to assess biomarker performance is using the non-parametric analysis receiver operator characteristic (ROC) curve (Xia *et al.*, 2013). When compared to a gold standard the ROC curve allows estimation of the diagnostic sensitivity and specificity of a newly proposed test. The sensitivity is defined as the probability a sample has been correctly classified as diseased and the specificity is the probability a sample has been correctly identified as non-diseased. The area under the ROC curve (AUROC) is a summary of all possible combinations of sensitivity and specificity that can be achieved, with the perfect test having a value of 1 (Greiner *et al.*, 2000). In this thesis, ROC analysis was performed using the online software tool Metaboanalyst (Xia *et al.*, 2012). To ensure a robust AUROC was generated, the process was repeated 100 times (cross-validation) and an average taken. For further validation of the ROC model, permutation testing was also performed. The null hypothesis of the permutation test was that "the optimal model found during the biomarker discovery process could also have been found if each sample had been randomly assigned a clinical outcome", as described by (Xia *et al.*, 2013). The original model performance was then statistically compared to the randomly assigned group and a p-value calculated. The lower the p-value the lower the chance the randomly assigned group performed at a similar level to the original set.

2.8.5 Determination of parasite burden

Strongyle burden was characterised for all animals in this thesis apart from horses sampled in the current chapter, section 2.9.2.1. However, these horses had received an anthelmintic (ivermectin) 4 weeks prior to sampling to target strongyles. Tapeworm (*Anoplocephala perfoliata*) burden was characterised by gold standard (counting of worms in the caecum) and for this reason diagnosis of *A. perfoliata* was only possible for animals sampled post-mortem in **Chapters 5** and **6**.

Strongyle burdens were determined by worm faecal egg counts (FEC) using a centrifugal floatation technique sensitive to one egg per gram (e.p.g.) (Christie & Jackson, 1982). The method was as follows: water (10 ml) was added to 10 g of faeces in a plastic beaker and stirred into a slurry with a spatula. Two subsamples of 5 ml each were pipetted through a sieve into a conical flask. A further 5 ml of water was pipetted through the sieve to rinse the

contents. The beaker was then agitated, and the liquid was poured into a centrifuge tube and spun for 2 minutes at 200 *g*. The supernatant was removed leaving 1 cm of liquid and an undisturbed pellet. Saturated salt solution (10 ml) was added and the centrifuge tube was slowly inverted to re-suspend the pellet before being centrifuged for a further 2 minutes at 200 *g*. The centrifuge tube was then clamped, and 1 cm of liquid poured into a cuvette. The cuvette was then left to stand for 5 minutes before reading under a microscope and all eggs were counted and recorded as eggs per gram (e.p.g.) For this thesis, faecal egg count (FEC) categories of high (500 e.p.g), medium (200-499 e.p.g) and low (0-199 e.p.g) were as outlined by American Association of Equine Practitioners (AAEP) and work by Nielsen and colleagues (Kaplan & Nielsen, 2010; Nielsen *et al.*, 2010a, 2013). Specifically for **Chapter 5** and **Chapter 6** two categories: low strongyle count (LSC) of 0-10 e.p.g and high strongyle count (HSC) of $\geq$200 e.p.g were used in addition to compare to (Peachey *et al.*, 2018).

2.9. Method development of the preparation steps of equine faecal samples for HS-SPME-GCMS

2.9.1 Introduction

The preparation steps of samples for metabolomics study are vital for achieving optimal results and consistent data in the later stages. Prior to this thesis there was no published method for the preparation of equine faecal samples for HS-SPME-GCMS VOC profiling; therefore, the aim of this work was to develop an optimal method. Previous work had established an optimal preparation method for HS-SPME-GCMS for murine and human faeces (Reade *et al.*, 2014). Furthermore, a standardised method would allow comparison of results between laboratories.

It has been shown that sample mass, vial volume and SPME fibre coating influences the numbers of VOCs and their abundance in human and murine faecal samples. In addition, the optimal preparation steps differed between human and murine faeces (Reade *et al.*, 2014). Therefore, it was proposed that a unique method would be required for optimal VOC profiling using GCMS for equine faeces.

2.9.2 Materials and Methods

2.9.2.1 Sample collection

Naturally voided faecal samples were collected from four horses immediately after natural defaecation in June 2015. Samples were frozen at -20^{o}C within 0.5-3 hours of collection and

were stored until analysis. All horses were housed on the same premises (Phillip Leverhulme Equine Hospital, Leahurst Campus, Neston) and apart from exceptional circumstances (Table 2.4), were maintained under the same regimen of living out at pasture all year round and a diet supplemented with hay during the winter. The horses were kept at rest and stabled for periods during the daytime to be handled by students. During this time horses had access to *ad libitum* hay. Details of collection location and animal demographics can be seen in Table 2.4. Parasite testing was not carried out as all horses had been treated for intestinal parasites with ivermectin and praziquantel 4 weeks previously.

Horse	Sex	Age (years)	Breed	Other details	Time of collection	Time from collection to freezing	Location of sample collection
L1	M	20	TBX	Dental problems – diet supplemented with commercial feed 16+ (Dodson and Horrel©) twice per day during the winter.	11:35	2 hrs	Stabled (since 8:30 am)
L2	M	18	ISHX		13:00	30 minutes	Stabled (since 8:30 am)
L3	G	13	ISHX	Currently stabled for foot abscess. Received oral phenylbutazone 48 hours prior to collection. *Ab libitum* diet of hay.	10:50	2.5 hrs	Stabled (for seven days)
L4	G	12	TBX		10:25	3 hrs	Paddock

Table 2.4 Horse demographics for sample preparation steps investigations. Key: M = mare, G = gelding (castrated male), TBX = Thoroughbred cross, ISHX = Irish sports horse cross.

2.9.2.2 Mass optimisation

Triplicates of 100 mg (mean and SEM) (105.5 ± 1.8 mg), 1000 mg (1003.0 ± 4.1 mg) and 2000 mg (2005.8 ± 5.1 mg) of faeces from each horse were divided into 10ml vials. The SPME fibre DVB-CAR-PDMS was used to extract VOCs before injection into the GC oven.

2.9.2.3 SPME fibre coating

Two different fibre coatings were tested to determine whether the type of fibre coating used in SPME-GCMS analysis influences the VOC profile. The SPME fibre coatings chosen were CAR-PDMS and DVB-CAR-PDMS. Masses of 1000 mg (1008.9 ± 3.3 mg) were placed into six vials for each horse. Three replicates were assigned to a CAR-PDMS fibre coating group and

3 replicates for a DVB-CAR-PDMS group for each horse. The samples then underwent HS-SPME-GC-MS analysis.

2.9.2.4 Vial volume

Vials of 10 ml and 20 ml were chosen to compare 1000 mg of faeces. Three 10 ml and three 20 ml vials containing 1000 mg (mean and SEM, 1003.9 ± 2.3 mg) of faeces from each horse underwent HS-SPME-GCMS using a DVB-CAR-PDMS SPME fibre to extract VOCs.

2.9.2.5 Technical replicates

A total of nine technical replicates of 1000 mg of faeces were analysed from each of the four horses. All samples were contained in 10 ml vials and VOCs were extracted using the SPME fibre DVB-CAR-PDMS. To assess variation across technical replicates, statistical analysis was performed.

2.9.2.6 Homogenisation

Aliquots of 1000 mg (mean and SEM, 1002.6 ± 0.45 mg) were placed in 10 ml vials from the exterior (n=3) and interior (n=3) of a faecal ball collected from each horse. Additional mixed material (n=3) was also divided in the same way. GCMS analysis using the SPME fibre DVB-CAR-PDMS was carried out.

2.9.2.7 GCMS conditions and data processing

GCMS conditions were as described in section 2.8.2. The data were processed, and appropriate statistical analysis applied as detailed in section 2.8.4.

2.9.3 Results

2.9.3.1 Mass optimisation

The mean (± SD) number of VOCs detected from each sample mass were 59 (± 9.5), 79 (± 8.8) and 80 (± 4.8) for 100 mg, 1000 mg and 2000 mg of sample, respectively. Significantly fewer VOCs were detected in 100 mg than 1000 or 2000 mg samples ($p<0.01$, ANOVA, Tukey's HSD test). There were 13 VOCs (Appendix 2.1) that were exclusive to 1000 mg and 2000 mg samples. All compounds detected were present in at least one sample of 1000 mg whereas three compounds (butanoic acid, 3-methyl-, propyl ester; propanoic acid, 2-methyl-; acetic acid) were missing from 2000 mg samples. Two compounds (2-decanone and 2-nonanone), were significantly lower in abundance in 100 mg than 1000 mg and 2000 mg samples ($p<0.01$,

ANOVA, Tukey's HSD test and FDR corrected). The CoV of VOC peak area was calculated for VOCs shared between replicates of each sample mass. Coefficient of variation values were 1.6-13%, 1.1-14.4%, 0.8-12.5% for 100 mg, 1000 mg and 2000 mg samples, respectively. A PCA is shown in Figure 2.2 which represents the VOC profiles of each sample mass. A full set of PC1 and PC2 scores are listed in Appendix 2.2.

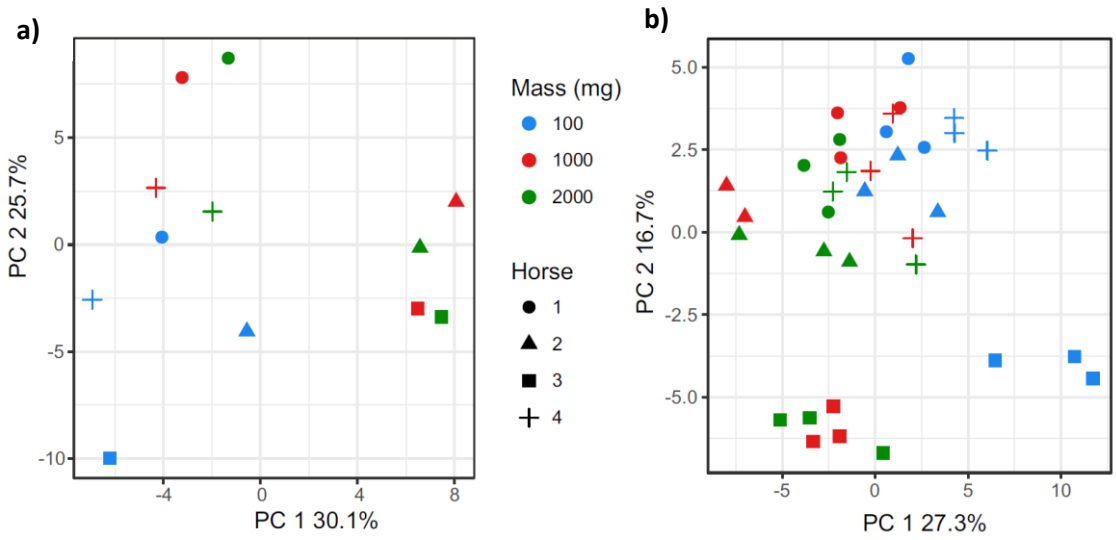


Figure 2.2 PCAs of the VOC profiles of 100, 1000 and 2000 mg samples of horse faeces analysed by HS-SPME-GCMS. In a) technical replicates were averaged, in b) technical replicates are shown.

2.9.3.2 SPME fibre coating

Mean (± SD) number of compounds detected by the CAR-PDMS SPME fibre was 52 (± 11.2) and DVB-CAR-PDMS was 78 (± 7.8). Significantly more VOCs were detected with a DVB-CAR-PDMS fibre than with a CAR-PDMS fibre ($p<0.01$, t-test). A total of 21 compounds were exclusive to the DVB-CAR-PDMS SPME fibre and one compound was exclusive to the CAR-PDMS SPME fibre. A list of compounds exclusive to each fibre is in Appendix 2.3. One compound (benzaldehyde) was significantly greater in abundance in samples exposed to the DVB-CAR-PDMS SPME fibre than the CAR-PDMS SPME fibre ($p<0.05$, ANOVA, Tukey's HSD test and BH corrected). A PCA is shown in Figure 2.3 which represents the VOC profiles of each fibre coating. A chromatogram overlay was generated for the DVB-CAR-PDMS and CAR-PDMS fibres for one of the horses (H2) and is shown in Figure 2.4.

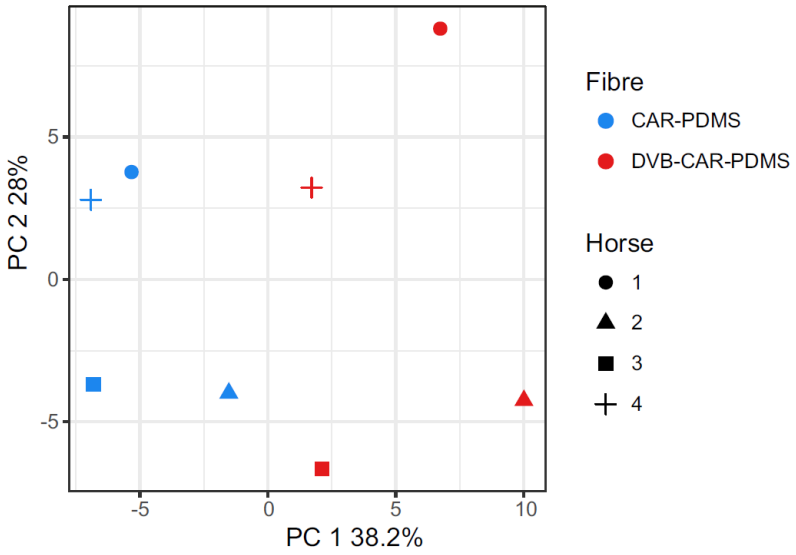
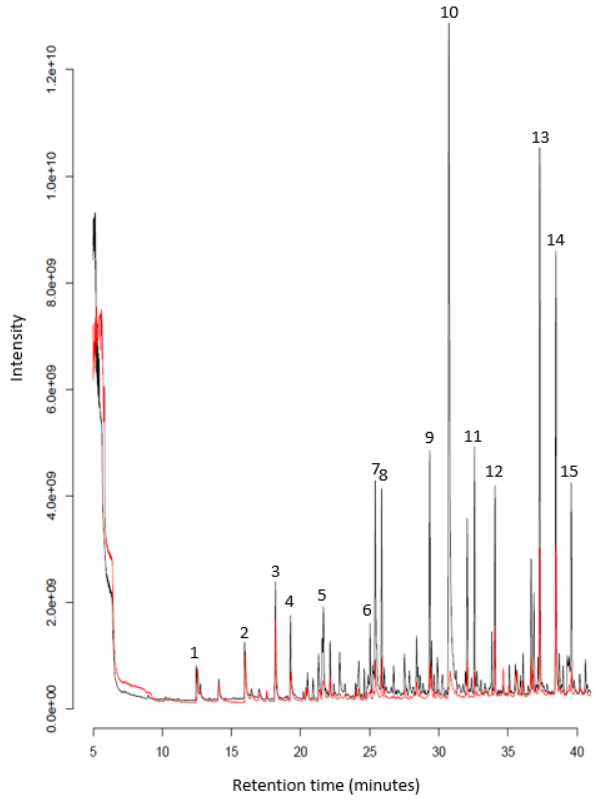


Figure 2.3 A PCA of the VOC profiles of VOCs extracted using DVB-CAR-PDMS and CAR-PDMS of horse faeces analysed by HS-SPME-GCMS. Three technical replicates were performed of each condition, per horse. Replicates were averaged prior to PCA.



Key:

1. Butanal, 3-methyl-
2. Propnaoic acid
3. Hexanal
4. Butanoic acid
5. Styrene
6. n-Caproic acid vinyl ester
7. 5-Hepten-2-one, 6-methyl
8. Octanal
9. Nonanal
10. P-Cresol
11. Decanal
12. 1-cyclohexene-1-carboxaldehyde, 2, 6-trimethyl-
13. Unclassified compound
14. Heptadecane, 2, 6, 10, 14-trimethyl-
15. Pentadecane

Figure 2.4 An overlay of chromatograms generated from the HS-SPME-GCMS analysis of faeces of horse 2 (H2). The black trace represents the DVB-CAR-PDMS fibre and the red is CAR-PDMS.

2.9.3.3 Vial volume

The mean (± SD) number of VOCs was 83 (± 9.6) and 78 (± 9.3) in 10 ml and 20 ml vials, respectively. Differences in mean VOC numbers between 10 ml and 20 ml vials were not statistically significant (p=0.48, t-test, FDR corrected). Two compounds (propanoic acid, 2-methyl- and acetic acid, methyl ester) were exclusive to 20 ml vials and four compounds (propanoic acid, 2-methyl-, methyl ester; 2-hexanone; propanoic acid, 2-methyl-, propyl ester; 1-nonanol) were detected from 10 ml vials only. There was no significant difference in the abundance of any compounds shared between 10 ml and 20 ml vials (p>0.05, t-test, FDR corrected). A PCA of the VOC profiles of the two vial volumes is shown in Figure 2.5.

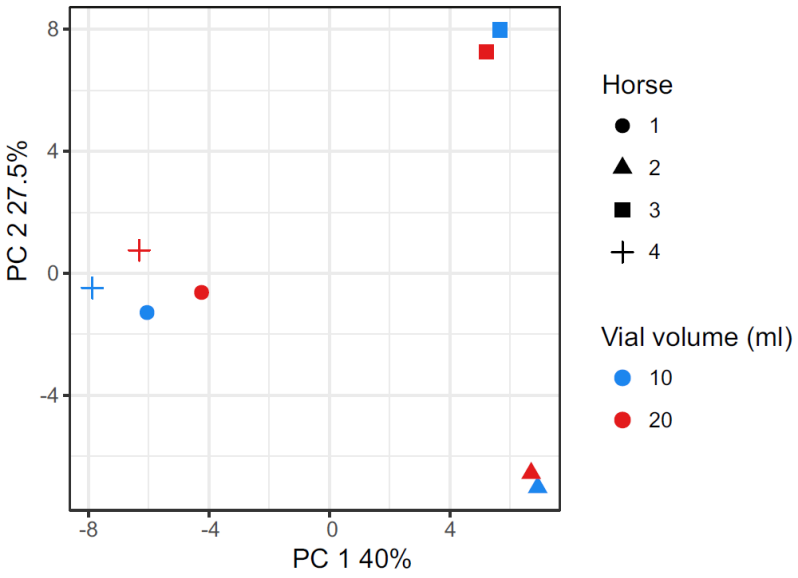


Figure 2.5 A PCA of the VOC profiles of horse faeces analysed by HS-SPME-GCMS in either 10 ml or 20 ml HS vials. Three technical replicates were performed of each condition, per horse. Replicates were averaged prior to PCA.

2.9.3.4 Technical replicates

The CoV was performed on the peak areas of VOCs shared across nine technical replicates of 1000 mg of faeces for each horse. The CoV for VOC peak area for H1 ranged between 0.7 and 8.6%, 1-11% for H2, 0.9-8.3% for H3 and 0.7 to 9.6% for H4. A PCA based on the VOC profiles of all technical replicates was constructed and is shown in Figure 2.6.

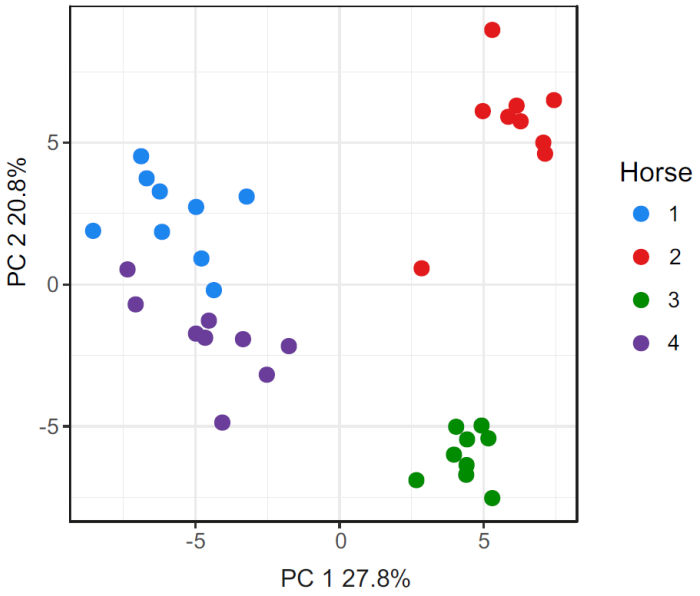


Figure 2.6 A PCA of the VOC profiles of the faeces of four horses analysed by HS-SPME-GCMS. Nine technical replicates were analysed per horse.

2.9.3.5 Homogenisation

Compounds did not significantly alter in abundance (ANOVA, Tukey's HSD test, FDR corrected) or in absence/ presence (Fishers' exact test) between the portions of faeces. However, there were three compounds exclusive to the outer portion of the faeces: benzene, propyl-; benzene, 1,2,4-trimethyl- and 3-pentanone, 2-methyl-. The number of VOCs did not significantly differ between portions of faeces (p= 0.9, ANOVA). A PCA showing how samples grouped for the portion of faeces and for individual horse is shown in Figure 2.7. A PERMANOVA analysis revealed that the portion of faeces (R^2 = 0.09, p=0.35) and individual horse factors (R^2= 0.19, p=0.07) were not able to describe any variation in the data. The CoV of shared VOC peak areas was calculated for each horse and the highest and lowest values are shown in Table 2.5.

Horse	VOC (CoV highest)	VOC (CoV lowest)
1	Toluene (15.3%)	1-Propanol, 2-methyl- (0.3%)
2	Propanoic acid, propyl ester (12.1%)	Propanal (0.2%)
3	Furan, 2-methyl- (8.2%)	Dodecane, 2,6,10-trimethyl- (0.2%)
4	Propanoic acid, propyl ester (10.4%)	2-Pentanone (0.1%)

Table 2.5 The highest and lowest coefficient of variation of shared VOC peak areas across the inside, outside and homogenised samples of the faeces of four horses.

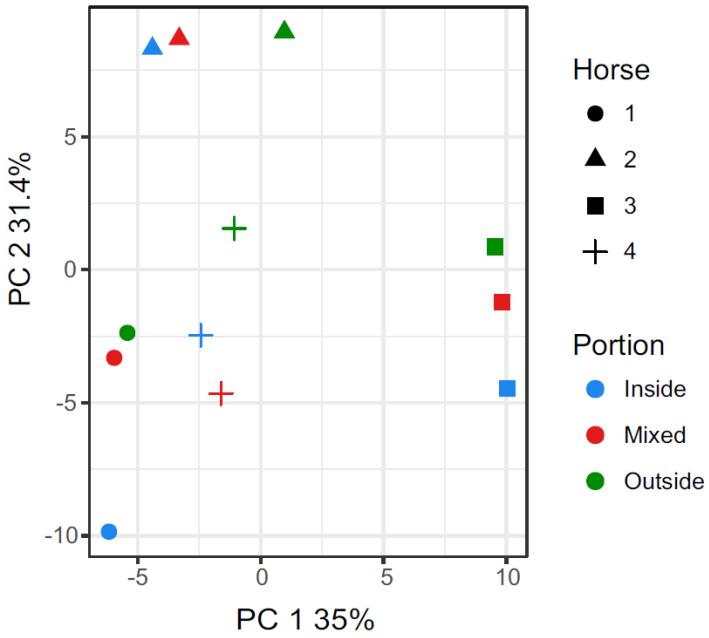


Figure 2.7 PCA homogenisation principal component analysis showing clustering for the portion of faeces and horse based on the abundance of compounds within each sample. Three technical replicates were performed of each condition, per horse. Replicates were averaged prior to PCA.

2.9.4 Discussion

2.9.4.1 Sample mass

A mass of 100 mg of faeces produced fewer VOCs than 1000 or 2000 mg. Masses of 1000 and 2000 mg showed very little variation between them and therefore 1000 mg may be considered an optimal mass for HS-SPME-GCMS of equine faeces. Masses greater than 2000 mg were not investigated as only minor differences were seen between 1000 and 2000 mg samples. It was likely that saturation of the SPME fibre or over-loading of the instrument had started to occur (Su Yin Ng *et al.*, 2012). The PCA plot (Figure 2.2) accounted for 55.8% of the variance in the data set. Pentadecane; undecane, 2,6-dimethyl- and tetradecane were among the VOCs responsible for most variation seen in PC1. For PC2, butanoic acid, 2-methylbutyl ester; methyl propionate and propanoic acid, ethyl ester were VOCs responsible for the most variation.

2.9.4.2 Fibre type

A greater number of VOCs were obtained when using DVB-CAR-PDMS fibre, rather than CAR-PDMS alone. The use of multiple fibre coatings has been shown to increase the diversity of VOCs obtained (Dixon et al., 2011; Reade et al., 2014). The choice of fibre coating is a crucial factor affecting SPME as there is not a single fibre coating suited to all analytes. A DVB coating

is mainly mesoporous (contains pores with diameters between 2 and 50 nm) , has a trapping range of C6-C15 and because of these properties is more suited to extracting medium and high molecular weight compounds (Mani, 1999; Gianelli *et al.*, 2002). Whereas a CAR coating is microporous (contains pores less than 2 nm in diameter), has a trapping range of C2-C12 and is more suited to low molecular weight compounds (Mani, 1999; Gianelli *et al.*, 2002). In contrast, a more marked variation in VOC diversity between CAR-PDMS and DVB-CAR-PDMS was observed in the present work than in human faeces (Couch *et al.,* 2013). Therefore, the use of DVB and CAR fibre coatings, combined with PDMS, appears to yield the best results for equine faeces. This is supported by the findings of Bianchi *et al*., when studying short chain fatty acids from an *in vitro* colonic fermentation model (Bianchi *et al.,* 2011). Therefore, a combination of the three coatings rather than two results in a net increase in diversity of compounds and offers a wider range of extraction.

2.9.4.3 Vial volume

The volume of the vial headspace containing 1000 mg of faeces did not influence the number of VOCs obtained from the headspace of faeces. Furthermore, neither the 10 ml nor 20 ml vial demonstrated a clear advantage over the other in terms of detecting VOCs at a higher abundance. These findings agree with work by others that an increase in headspace volume does not have an impact on analyte detection (Cho *et al.*, 2003). However it may depend on sample matrix, as found in murine faeces and human faeces (Reade *et al.*, 2014). A smaller headspace volume resulted in a higher yield of VOCs for murine samples, whereas for human faeces (and horse faeces in the present work) an alteration to the headspace volume had negligible impact. The SPME theory suggests that decreasing the volume of headspace increases the chance of compounds being detected. A vial volume increase from 10 ml to 20 ml may not have provided a sufficient increase in headspace for a difference in VOCs to be seen in the present work. It can be concluded that 10 ml or 20 ml vials are suitable for HS-SPME-GCMS analysis of 1000 mg horse faeces with insignificant impact on the presence or abundance of VOCs obtained.

2.9.4.4 Technical replicates

Analysing technical replicates and taking an average is a common practice in metabolomics (Bader *et al.*, 2016). Within each horse the CoV of shared VOC peak areas were below 11%, which is less than 30%, the threshold considered to be acceptable for biological samples (Drabovich *et al.*, 2013). Here it has been shown that up to nine technical replicates are able to show reproducible results. For the rest of this thesis a reasonable number (in terms of time

and cost) of three technical replicates per sample has been chosen to minimise random errors.

2.9.4.5 Homogenisation

Differences in VOC numbers and abundance between regions of the faecal ball were not observed. Three compounds were unique to the sample exterior: benzene, propyl-; benzene, 1,2,4-trimethyl- and 3-pentanone. It is likely that these are contaminants from the plastic bag used to pick up and store the samples as all three compounds have been reported to be emitted from plastics (Wypych, 2017). A PCA investigating the abundance of compounds found within each replicate did not demonstrate distinct clustering of the inside, mixed and outside of the faeces (Figure 2.7). These data suggest that there is no unique VOC profile to different areas of the faecal ball.

Clustering of individual horses was apparent (Figure 2.7) and according to PERMANOVA analysis horse accounted for 19% of variation in the data. However, the p-value (p=0.07) reported was not significant but indicates a trend that individual horse was a contributing factor. In addition, it is worth noting that a sample size of four horses may not have provided enough statistical power to detect any differences. In human stool samples with five participants, clustering for individuals rather than the portion of stool sampled was also observed (Gratton *et al.*, 2016). In PCAs constructed for each participant the latter study found some separations between portions of stool, but patterns were not consistent between individuals. Similar results were seen in this work with individual horses (Appendix 2.4). In further support of this, the shared VOC peak areas that had the highest and lowest CoV were predominantly individual specific (Table 2.5). The exception was propanoic acid, propyl ester which was the VOC with the highest CoV for both Horse 1 and Horse 4. Interestingly the samples for Horse 1 and Horse 4 clustered closely together, indicating a similarity between the VOC profiles of these two horses (Figure 2.7). In work investigating the microbiota of an equine faecal ball sampled on the surface and homogenised there were no significant differences (Beckers *et al.*, 2017). However, authors in the latter study stated the outer region was more variable than the homogenised sample and would still recommend homogenisation as standard.

2.9.4.6 Extraction time and temperature

Extraction time and temperature are considered to be important factors influencing the VOCs extracted from faeces (Couch *et al.*, 2013). However, these factors were not investigated

here as previous work in both murine and human faeces observed that an extraction time of less than 20 minutes yields fewer VOCs and extraction durations for longer than 20 minutes do not yield greater VOCs, indicating the fibre has reached equilibrium (Reade *et al.*, 2014). In the same work temperatures below 60°C resulted in fewer VOCs being detected and a temperature of 70°C did not increase the numbers of VOCs detected. The use of SPME to extract VOCs has been applied to horse faeces previously. However the PDMS-DVB fibre was exposed to samples that had been freshly defrosted and were not heated to 60° as in the current work (Stavert *et al.*, 2014). In the work by Stavert *et al.*, (2014) 7 volatiles were obtained from a single sample of 10 g. This is compared to an average of 63 (SD ± 8.4) VOCs obtained from 1 g of faeces from four horses in the present investigation. Although Stavert *et al.,* (2014) exposed the fibre for 30 minutes they did not heat the samples to 60°C, indicating that a lower extraction temperature yields fewer VOCs. Other factors including column material, run time and data processing methods could also have contributed to the differences seen between these two laboratories. Others have shown extraction times of 18 hours extract considerably greater VOCs from human faeces (Couch *et al.*, 2013). However, such extraction times are impractical for high-throughput analysis and therefore would not be an appropriate addition to this method. Furthermore an extraction duration of 20 minutes and a temperature of 60°C has been successfully applied for poultry, mice and human faecal samples therefore it is likely this will not differ for horse faeces (Garner *et al.*, 2007, 2008; Reade *et al.*, 2014).

2.10 Pilot studies of sample storage methods affecting equine faecal VOC profiles

2.10.1 Introduction

The processing of fresh samples is not always practical and therefore some form of sample storage is usually required. Knowledge of the effects of freezing temperatures, durations of storage and the time window from collection to getting samples into storage on faecal VOCs is highly important for experimental design and for generating repeatable and reliable results.

A time lapse of up to 12 hours prior to freezing is known to have an effect on the microbiome of equine faecal samples but the effect on VOCs has not previously been investigated (Beckers *et al.*, 2017). Exposure of samples to ambient air may result in VOCs being lost from the sample prior to analysis. It is not known precisely when an adverse effect on VOCs prior to the freezing of equine faeces occurs. The effect of storage conditions has been found to

strongly influence the human faecal VOC profile (Saric *et al.*, 2008; Berkhout *et al.*, 2016; Gratton *et al.*, 2016). The effect of storage conditions and duration of storage on equine faecal VOCs is yet to be investigated.

The aims of the following storage investigations are: (1) to investigate whether prolonged exposure for up to 24 hours in the collection environment prior to freezing has an impact on VOC yield and (2) to investigate the impact of long term -20°C and -80°C storage on the equine faecal VOC profile. The impact of long-term storage on faecal VOCs is a key factor to consider for the longitudinal studies carried out later in this thesis. As with sample preparation steps, standardisation of storage methods would allow more accurate comparisons to be made between different experiments and different laboratories.

2.10.2 Methods

2.10.2.1 Horses

Faecal samples were collected after spontaneous voiding of faeces from a mixed-breed pony mare (P1) in March 2015 (Part A), May 2016 (Part B), January 2017 (Part C) and April 2017 (Part D). The pony, (date of birth: June/2000) was maintained at pasture with *ad lib* access to hay and water. A dose of moxidectin was administered in January each year. Faecal egg counts (0 e.p.g) were recorded each year in mid-April and at this time Praziquantel (2.5 mg /kg bodyweight) was administered. Extraction of VOCs from samples was performed as detailed in section 2.8.1 using the SPME fibre DVB-CAR-PDMS and samples were contained in 10 ml headspace vials. GCMS conditions were as described in section 2.8.2.

Part A) Time from collection to freezing and freeze-thawing

A faecal sample was collected from P1 in March 2015. The sample was mixed manually and a portion of the faeces (approximately 50 g) was then stored in a freezer at -20°C after 1, 2, 4, 8, 12 and 24 hours of environmental exposure. Samples were defrosted at room temperature and 1000 mg (mean and SEM, 1005.18 $\pm$ 0.85 mg) divided into vials, with three replicates of each condition. Nine additional aliquots were also divided into HS vials and were frozen at -20°C. Three samples were defrosted once, three samples were defrosted twice, and three samples were defrosted three times before HS-SPME-GCMS. For cycles 2 and 3 samples were subject to a one hour defrost at room temperature before being returned to the freezer (-20°C).

Part B) The effect of long-term storage on faecal VOCs at -20°C and -80°C

In May 2016 a faecal sample was collected from P1 to investigate the effect of storage on faecal VOCs. The sample was immediately transported to the laboratory (< 2 hours). A sample weight of 1000 mg (mean and SEM, 1006.18 ± 0.60 mg) was divided into headspace vials and was stored at either -20°C (n=12) or -80°C (n=12) until analysis either one week, 6 months or 12 months later. Three fresh aliquots of the sample underwent HS-SPME-GCMS using SPME fibre DVB-CAR-PDMS within 2 hours of defecation to compare these with the frozen samples.

Part C) The effect of snap freezing and storage in 15 ml Falcon tubes

A faecal sample was collected from P1 in January 2017. The sample was transported to the laboratory as detailed in Part B). On arrival at the laboratory the sample was homogenised and divided into 15 ml Falcon tubes (approximately 5 g of sample in each). Several tubes were snap frozen and were placed in a -20°C freezer (n=3) or in a -80°C freezer (n=3). At the same time three tubes were placed in a -20°C freezer and three in a -80°C having not been snap frozen in liquid nitrogen. Falcon tubes (volume of 15 ml) were chosen as glass HS vials could not be immersed in liquid nitrogen. The samples were stored in the Falcon tubes and were partially thawed at room temperature prior to analysis and 1000 mg (mean and SEM, 1000.6 ± 0.75 mg) was divided into headspace (HS) vials one week later. Three aliquots of fresh sample were aliquoted into HS vials and loaded onto the GCMS two hours post collection.

Part D) The effect of snap freezing and storage in cryovials

In April 2017 a faecal sample was collected from P1 and was transported to the laboratory as detailed in Part B). On arrival the sample was homogenised and divided into cryovials (2 ml). Cryovials were then treated as follows: snap frozen and stored at -20°C (n=3) or -80°C (n=3), stored at -20°C (n=3) and -80°C (n=3) but not snap frozen beforehand. Aliquots (1000 mg) of fresh samples (n=3) were placed in HS vials and were loaded onto the GCMS two hours post collection. A sample of 1000 mg (mean and SEM, 1000.7 ± 3.28 mg) was also placed in a HS vial and stored at -20°C to allow comparison of cryovial with HS storage.

2.10.2.2 Data processing

Data were processed as described in section 2.8.4. The use of one animal in this investigation did not allow statistical analysis to be applied to the data. PCA plots for visual comparison, stacked plots for chemical classification comparison, changes in VOC numbers and CoV between shared VOC peak areas were assessed.

2.10.3 Results

Part A) Time from collection to freezing and freeze-thawing

The weather conditions at time of collection and the 24 hours samples spent exposed to outside air (to replicate normal sampling conditions) before freezing were recorded based on observations taken from the nearest weather station (Rhyl) to the sampling location. Air temperatures reached a low of 3.1^{o}C and a high of 9.2^{o}C, mean hourly wind speed ranged between 3 and 8 knots and rainfall was at 0 mm.

The mean numbers of VOCs for each condition and a boxplot showing the spread of the data is shown in Figure 2.8a. The highest number of compounds shared with the sample frozen after 1 hour was 4 hours (98%) followed by 2 hours (97%), 8 hours (95%), 12 hours (94%) and 24 hours (90%). A PCA based on the VOC profiles of the 6 conditions is shown in Figure 2.8b. The loading data for the 15 highest scoring compounds for PC1 and PC2 of the PCA (Figure 2.8b) are listed in Table 2.6. A full list of scores for PC1 and PC2 is in Appendix 2.5. Box and whisker plots of a selection of compounds with the highest scoring compounds were constructed to show the pattern of change over time (Figure 2.9). The CoV of shared peak areas across all conditions were calculated; the lowest value was for pentadecane (0.6%) and the highest for 1-nonanol (9.6%).

The mean (± SD) number of VOCs after each freeze thaw cycle were 83 (± 1.2), 84 (± 1.2) and 84 (± 3) after the first, second and third cycles, respectively. Over the three cycles 96% of compounds were shared. Compounds that altered in their presence in samples were: octane and 2-heptanal (Z)- and isopropyl alcohol. Octane and 2-heptanal were present in cycles 2 and 3 only and isopropyl alcohol was no longer detected after 2 cycles. The CoV of shared VOC peak areas across the three cycles ranged between 0.5 and 5.4%. A PCA was constructed based on the VOC profiles of the three freeze thaw cycles and is shown in Figure 2.10.

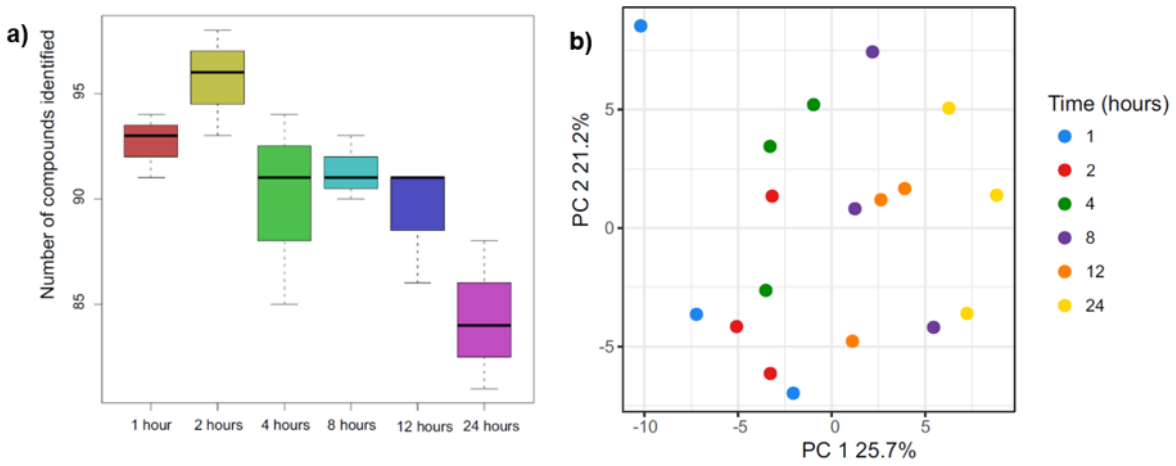


Figure 2.8 a) Box and whisker plot of compound numbers and b) A PCA of the VOC profiles of equine faeces frozen after 1, 2, 4, 8, 12 and 24 hours exposure to collection environment.

Compound	PC1	Compound	PC2
Octanal	0.18	Nonane	-0.14
Furan, 2-pentyl-	0.17	Toluene	-0.15
1-Propanone, 1-cyclopropyl-	0.17	Propanoic acid, 2-methyl-, methyl ester	-0.15
Furan, 2-ethyl-	0.17	Benzene, 1,3-dimethyl-	-0.15
Nonanal	0.17	2,3-Butanedione	-0.15
Heptanal	0.16	Pentadecane	-0.16
Furan, 2-methyl-	0.16	Heptadecane, 2,6,10,14-tetramethyl-	-0.16
2-Pentanone	0.15	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.16
5-Hepten-2-one, 6-methyl-	0.15	Dodecane, 2,6,10-trimethyl-	-0.16
2-Hexenal	0.15	1,7-Octadiene, 2,7-dimethyl-	-0.16
Benzaldehyde	0.15	Cyclohexanone, 2,2,6-trimethyl-	-0.16
Decanal	0.15	Decane	-0.17
Hexanal	0.15	Tridecane	-0.18
2-Nonanone	0.15	Tetradecane	-0.18
Propanoic acid, propyl ester	-0.15	Octane	-0.19

Table 2.6 Scores (top 15) for PC1 and PC2 of the PCA in Figure 2.8b.

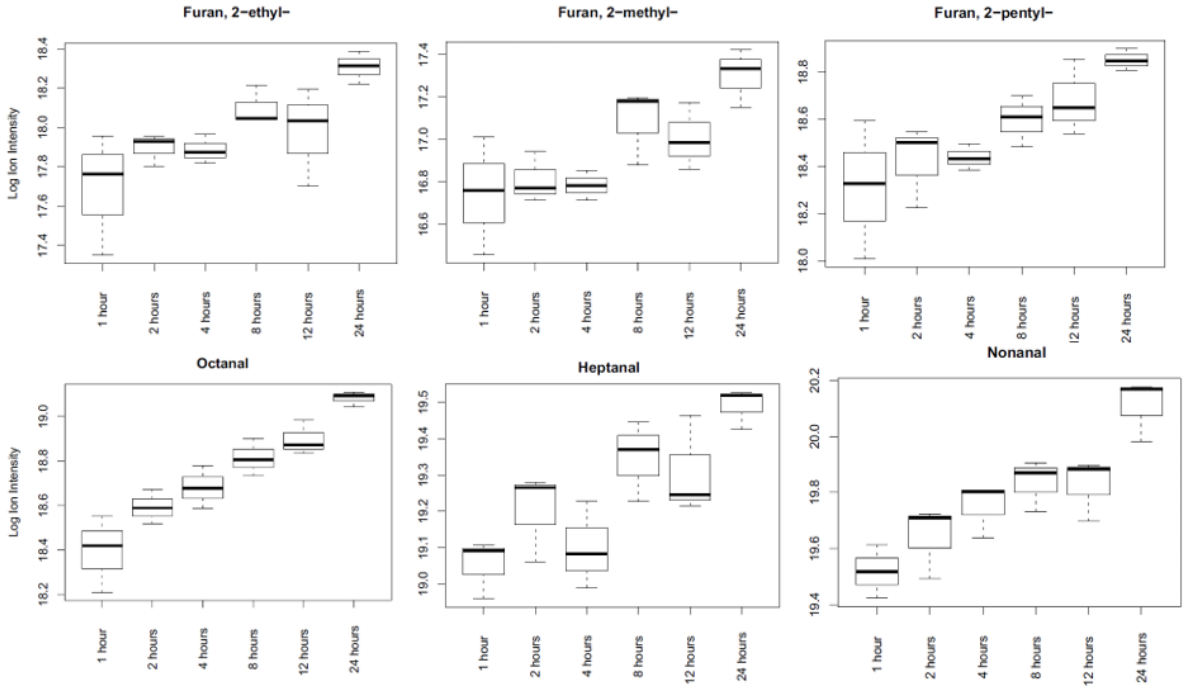


Figure 2.9 Box and whisker plots of a selection of furan and aldehyde compounds which had some of the highest scoring principal components in the PCA analysis (Figure 2.8b).

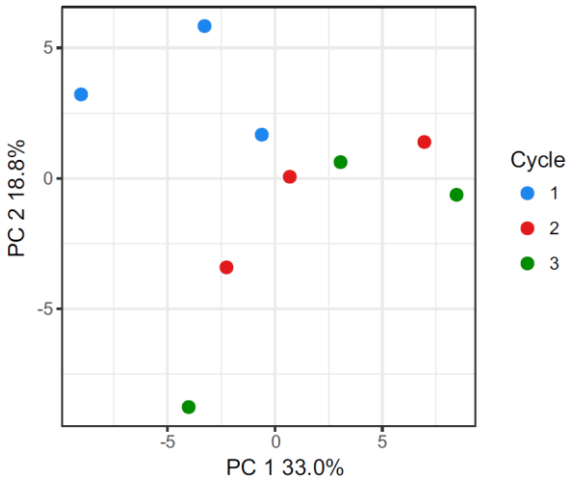


Figure 2.10 A PCA showing equine faecal samples subject to 1, 2 and 3 freeze-thaw cycles before HS-SPME-GCMS.

Part B) The effect of long-term storage on faecal VOCs at -20°C and -80°C

The mean number of VOCs identified in the fresh faecal sample, storage after 1 week, 6 months and 12 months at -20°C and -80°C are shown in Table 2.7. The percentages of VOCs shared between fresh and stored samples are also shown in Table 2.7. A PCA based on the

VOC profiles of the 7 storage conditions was constructed and is shown in Figure 2.11. A stack plot of the chemical classes of compounds found in each condition are shown in Figure 2.12

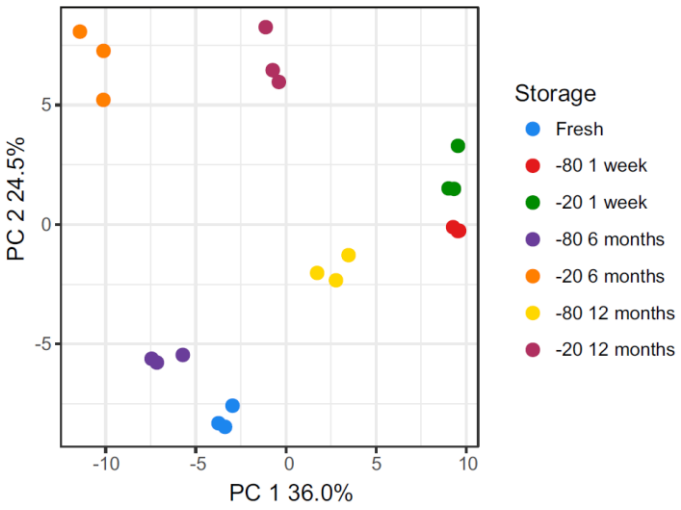


Figure 2.11 A PCA of equine faecal compounds detected in a fresh sample and samples stored at -20°C and -80°C for 1 week, 6 months and 12 months.

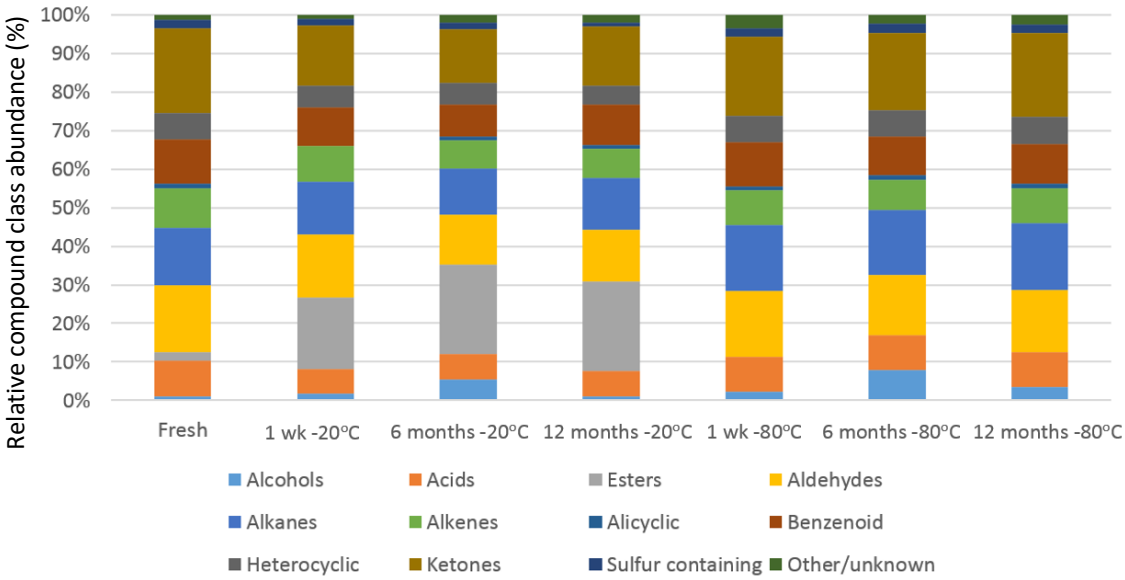


Figure 2.12 A Stacked plot of the chemical classes of equine faecal compounds detected in a fresh sample and samples stored at -20°C and -80°C for 1 week, 6 months and 12 months. Some key VOCs of various chemical classes were plotted in Appendix 2.6 to show their variability.

Storage conditions (°C)	Mean (± SD) numbers of VOCs
Fresh	79 (± 1.5)
1 week at -20	99 (± 1.5)
1 week at -80	82 (± 1.5)
6 months at -20	95 (± 3.6)
6 months at -80	80 (± 0.6)
12 months at -20	93 (± 2.5)
12 months at -80	80 (± 2.6)
Storage conditions (°C)	**Percentage (%) of VOCs shared**
Fresh and 1 week at -20	79
Fresh and 6 months at -20	65
Fresh and 12 months at -20	64
Fresh and 1 week at -80	91
Fresh and 6 months at -80	84
Fresh and 12 months at -80	84

Table 2.7 A table of the mean number of VOCs and percentages of VOCs shared between an equine faecal sample. Analysis by HS-SPME-GCMS two hours post collection, after storage for 1 week at -20°C and -80°C and after 6 months at -20°C and -80°C.

Part C) The effect of snap freezing and Part D) The effect of cryovial storage and snap freezing

The mean number of VOCs identified in the fresh faecal sample, storage after -80°C and -20°C snap frozen, -20°C and -80°C not snap frozen are shown in Table 2.8. The percentages of VOCs shared between fresh and stored samples are also shown in Table 2.8. For Part D) a comparison between the cryovial samples and HS vial is also demonstrated in Table 2.8. A PCA of the VOC profiles for each set of samples is shown in Figure 2.13. Stacked plots of the chemical classes of compounds found in each condition for both sets of samples are shown in Fig 2.14.

Falcon tubes (15 ml)			Cryovials (2 ml)		
Condition	**Percentage (%) of VOCs shared with a fresh sample**	**Mean number of VOCs (± SD)**	**Condition**	**Percentage (%) of VOCs shared with a fresh sample**	**Mean number of VOCs (± SD)**
-20°C not snap	79	89 (±4)	-20°C cryovial not snap	89	81 (± 0.58)
-20°C snap frozen	78	89 (±4.6)	-20°C cryovial snap	88	81 (0)
-80°C not snap	91	76 (±1.5)	-80°C cryovial not snap	94	80 (± 2.3)
-80°C snap frozen	94	75 (±1)	-80°C cryovial snap	97	79 (± 2.5)
Fresh		76 (±1.15)	Fresh		80 (± 2.6)
			-20°C HS vial	75	94 (NA)

Table 2.8 A table of the mean number of VOCs and percentages of VOCs shared between equine faecal samples that were fresh and stored in various conditions.

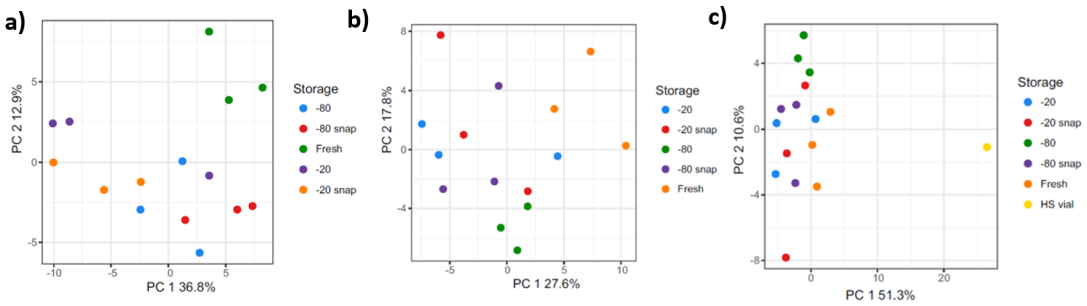


Figure 2.13 PCAs of equine faecal samples. Samples were **a)** frozen in 15 ml Falcon tubes **b)** frozen in cryovials **c)** same samples in b but with a 10 ml headspace (HS) vial included for comparison. All samples were analysed either fresh or were stored at -20°C or -80°C and were either snap frozen in liquid nitrogen or placed directly in a freezer without being snap frozen.

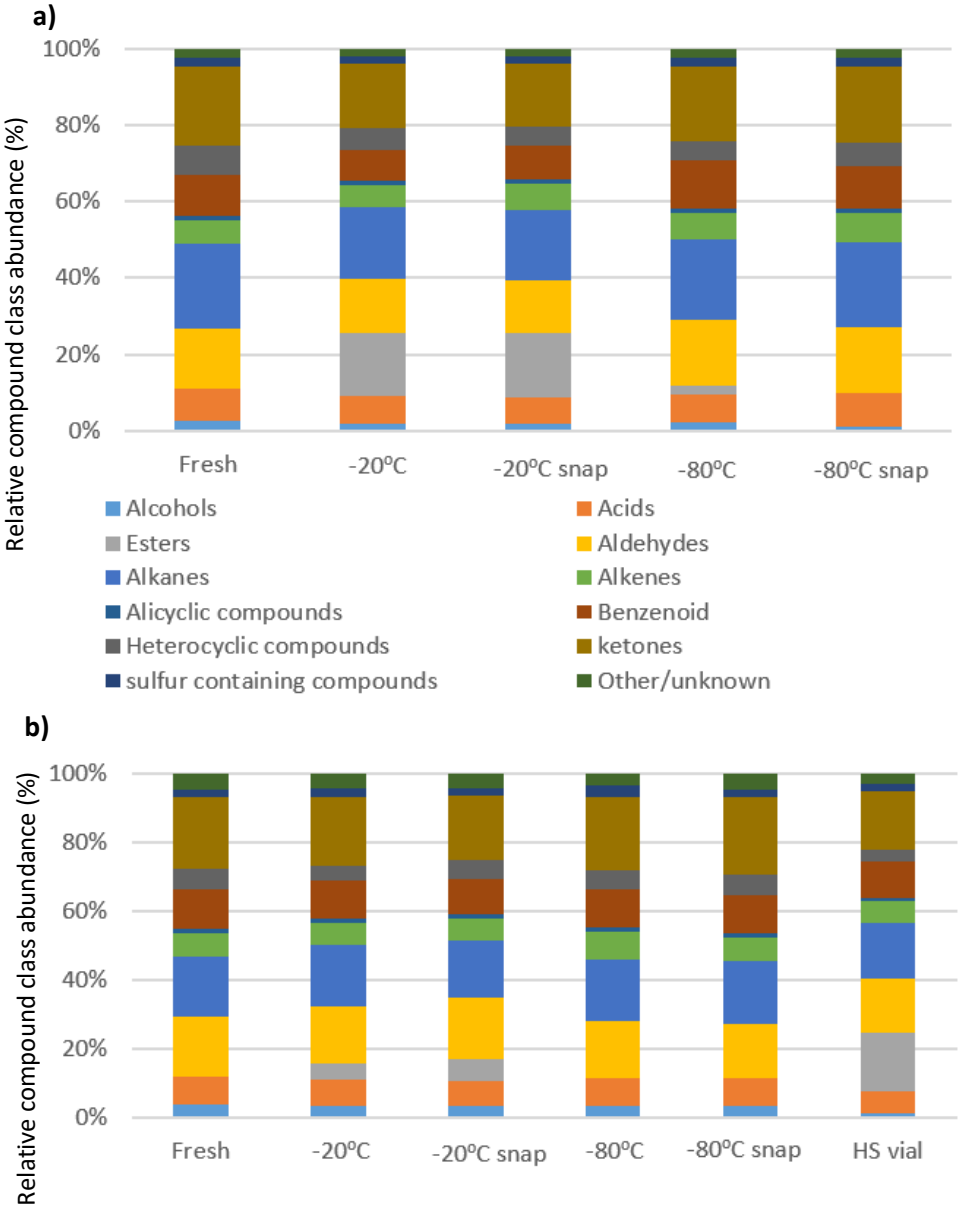


Figure 2.14 Stacked plots of the chemical classes of compound found in equine faecal sample. In **a)** samples were stored in 15 ml Falcon tubes and in **b)** samples were stored in cryovials (a 10 ml headspace (HS) vial stored at -20°C was included for comparison). All samples were analysed either fresh or were stored at -20°C or -80°C and were either snap frozen in liquid nitrogen or placed directly in a freezer without being snap frozen. Some key VOCs of various chemical classes were plotted in Appendix 2.7 to show their variability.

2.10.4 Discussion

Part A) Time from collection to freezing and freeze thawing

A total of 98% of VOCs were present in both the 1 hour and 4-hour samples. From 8 hours onwards, the numbers of VOCs shared with a 1-hour frozen sample decreased, with the fewest VOCs shared at 24 hours. The loss of VOCs over time may be because compounds are being lost to the ambient air and are not being replenished in the faeces as the anaerobic bacteria diminish. Previous work in human and bovine faeces revealed that aerobic exposure over the course of several days resulted in an overall decrease of VOCs (Laor *et al.*, 2007; Lin *et al.*, 2013). In pig slurry, branched chain fatty acids increased after 6 weeks storage at room temperature, possibly because of an alteration in microbial populations over this time and the establishment of aerobic bacteria (Hwang *et al.*, 2016).

A trend across PC1 shows some clustering in order of collection from 1 hour towards 24 hours (Figure 2.8a). Aldehydes were among the top compounds responsible for driving the most separation in PC1. Box and whisker plots demonstrated a linear increase in the abundance of aldehydes over time (Figure 2.9). The total amount of variation described in PC1 and PC2 is relatively low (46.9%). This is unsurprising as at 24 hours, samples still shared 90% of compounds with 1-hour frozen samples. Furthermore, the CoV of shared VOC peak areas were all below 10% indicating very little change in peak areas. Interestingly for PC2, alkanes were the main drivers of separation and the samples spread across PC2 were technical replicates. This suggests that alkanes are responsible for much of the variation between technical replicates; however, the total variation was still low for PC2 at 21.2%.

In work studying the impact of exposure to ambient air on the bacteria of equine faeces, it was found that alpha diversity increased initially followed by a subsequent decrease over time (Beckers *et al.*, 2017). In comparison to the present work, an initial increase in VOCs was observed in the latter study, followed by a decrease in VOC numbers over time. In a study characterising VOCs emitted from anaerobic and aerobic incubated soil, greater numbers of VOCs (27) were detected in anaerobic samples than aerobic samples (13) (Wheatley *et al.*, 1996). A number of studies examining the impact of ambient air exposure to the bacteria in faeces have reported a shift from anaerobic to aerobic species after 12 hours or more (Roesch *et al.*, 2009; Wong *et al.*, 2016; Beckers *et al.*, 2017). In relation to VOCs, Wheatley *et al*., found that the addition of ammonium sulfate (fertilizer) to soil in aerobic conditions significantly increased the proportion of aldehydes in aerobic but not anaerobic conditions. The evidence suggests that the aldehyde patterns seen in the current work (Figure 2.9) may

be indicative of a shift in bacterial populations from largely anaerobic (representative of the gut microbiota) to aerobic species over time. The link with aldehydes and aerobic bacteria in the current work can only be speculated upon as microbial populations were not characterised in this investigation.

The average air temperature recorded in the current work during sample exposure was 6.1°C, in comparison to the work by Beckers *et al*., where the average air temperature was 32°C. An increase in air temperature may influence the volatility of compounds as well as bacterial growth. This could offer some explanation as to why differences seen in the bacterial populations in the latter study were more pronounced than the subtle VOC results observed here. However bacterial sequencing and VOC profiling on the same set of samples is necessary to confirm these hypotheses.

The current investigation has provided an initial insight into whether equine faecal VOCs are affected when exposed to ambient air. Future work should involve sampling at separate times of the year to account for climatic changes and it would be of interest to combine omics – sampling the microbiota and VOC metabolome of the same samples – as VOCs in the faeces are often studied as a proxy for the microbiota. Beckers *et al.,* took samples from three horses and after 12 hours samples clustered by time rather than individual horse. It is currently unknown whether equine faecal samples would cluster in the same way according to VOCs, as samples were collected from a single horse in the current investigation. To confirm findings of this work, sampling multiple horses is required to allow statistical comparisons to be made and to determine a cut-off point for freezing samples post collection.

The freeze-thaw experiment conducted here was useful for the work to be carried out in this thesis as most samples will be stored at -20°C (longitudinal samples will be stored at -80°C during collection and transferred to -20°C on dry ice once all samples have been collected). As the samples were likely to be defrosted up to one or two times for the work to be carried out, it was important to establish to what extent defrosting and re-freezing at -20°C would have on the VOC profile. A separation of cycle 1 away from cycle 2 and 3 is shown in Figure 2.10, but PC1 represents just 33% of the variation in the plot. This is further supported that across the three cycles, samples shared 96% of VOCs and all shared VOC peak areas that had a CoV of less than 6%. These results indicate that the defrosting of samples up to three times has a minimal effect. In work by Gratton *et al.,* (2016), three cycles of freeze thawing of human faecal water also had minimal effect on SCFAs, but amino acids increased. The amino acids were non-volatile, so they would not be identified in the current work. In terms of

bacteria in faeces, freeze-thawing up to four times also had minor effects on the taxa present (Gorzelak *et al.*, 2015). However, after five cycles Bacteroidetes increased and Enterobacteriaceae decreased. Freeze thawing of human urine samples three times over 6 months demonstrated a much larger effect on the observed VOCs than the faecal VOCs in the present work (Semren *et al.*, 2018). The effect of freeze-thawing on faeces of human or animal species over 6 months or longer has not been investigated (Karu *et al.*, 2018). Time between freeze-thaw cycles may be an important factor in the overall effect of freeze-thawing. Therefore, an increase in the number of cycles or the length of time over which the freeze thawing occurred may have a greater impact on the faecal VOCs than observed in the current work. Further investigation should involve the inclusion of additional freeze-thaw cycles and at other temperatures (e.g. -80°C) on the same sample to control for horse variability.

Part B) The effect of long-term storage on faecal VOCs at -20°C and -80°C

The effect of storage length and temperature on faecal VOCs has received little attention (Deda *et al.*, 2015). In the present work a greater number of compounds were found in samples stored at -20°C than fresh or -80°C. An increase in compounds after one month of freezing (-80°C) has been observed by others (Saric *et al.*, 2008). Chemical classification of VOCs in the current study, revealed a substantial proportion of compounds in -20°C stored samples were esters. Fewer esters were seen in fresh samples and those stored at -80°C, indicating that the formation of esters is specific to -20°C storage. From this work it can only be speculated as to why esters were specific to storage at -20°C, possibly because at -20°C the sample may take longer to freeze and therefore esterification of acids continues compared to a sample frozen at -80°C. This phenomenon may be pony- or storage vessel-specific and requires further investigation. In human faeces it was observed that after 24 hours of storage at -20°C there was no difference in VOC profile between that and a fresh sample (Gratton *et al.*, 2016). Species differences are evident in preparation steps for metabolomics analysis (Reade *et al.*, 2014). It is likely these differences are because of the differing VOC profiles between species, largely attributed to diet and intestinal microbiota. Unique storage conditions may be required for diverse types of sample. It was interesting to note that the percentage of VOCs shared between the fresh sample and samples stored at -80°C for 6 and 12 months decreased. An example includes benzene, (1-methylethyl)- which was present in a fresh sample and a sample stored at -80°C for 1 week but was not present in any -80°C samples after 6 or 12 months. In contrast, 1-octen-3-ol was not present in the fresh sample but was detected in -80°C samples after storage of 1 week, 6 months and 12

months. From this work it is difficult to determine why VOCs were lost or gained after storage as we still know very little of the impact of freezing on VOCs (Berkhout *et al.*, 2016). It can be speculated that the loss or gain of VOCs may be attributed to the sample container (Mochalski *et al.*, 2009) or the effect of freeze-thawing on microorganisms present in faeces (Achá *et al.*, 2005).

The effect of snap freezing and cryovial storage (parts C and D)

Esters formed in the 15 ml Falcon tube samples after -20°C regardless of whether they were snap frozen or not (Figure 2.14). In the cryovial stored samples esters also formed at -20°C, but to a lesser extent (Figure 2.14). A 15 ml Falcon tube sample stored at -20°C (not snap frozen) shared 10% fewer compounds with a fresh sample, than the equivalent in a cryovial. For -80°C stored samples, snap freezing did not have an impact on chemical classes for either 15 ml Falcon tubes or cryovial stored samples.

The most compounds shared with a fresh sample were when snap frozen in a cryovial then stored at -80°C. The fewest number of compounds shared with a fresh sample was a 15 ml Falcon tube snap frozen then stored at -20°C. Visual separation of the fresh sample from the frozen samples was to a lesser degree for the cryovials than the 15 ml Falcon tubes as explained by PC1 (Figure 2.13). PC1 explained 27.6% of variation for the cryovials and 36.8% of variation for the 15 ml Falcon tubes. Therefore, cryovial storage at -80°C appeared to maintain original sample integrity the most.

Cryovials are specifically designed to hold biological specimens at very low temperatures which may explain why they performed best. Furthermore, the HS vials and 15 ml Falcon tubes were not filled with each sample as the cryovials were and the air space in the vessels may have contributed towards a greater formation of esters at the lower temperature. However, 15 ml Falcon tubes and HS vials were also effective at preventing ester formation at -80°C. Therefore, it can be concluded that storage at -80°C is most effective at maintaining original sample integrity. This is further improved when a sample is also stored in a cryovial at -80°C. Snap freezing of samples in liquid nitrogen prior to placing samples in a -20°C or -80°C freezer had very little effect regardless of the type of vessel used.

A HS vial was included in Part D) to demonstrate that the effect of storage in this vessel is consistent with previous findings and that this method of storage resembled a fresh sample the least (Part B). It would have been interesting to investigate -80°C HS compared to

cryovials as well as -20°C. Furthermore, the use of technical replicates here (e.g. n=3 instead of n=1) would have resulted in a more robust comparison.

2.10.4.1 Summary and limitations of the storage investigations

Over the course of 24 hours of a sample left in field collection conditions demonstrated a decrease in overall VOC numbers but an increase in aldehyde intensities were observed. The observed changes may be related to a shift from anaerobic to aerobic bacteria, but characterisation of the bacterial populations and VOCs in synchrony is required to confirm this theory. Furthermore, to understand the loss of VOCs in relation to their volatility the calculation of Henry's Law Constant (the concentration of a gas in the headspace above a liquid) could have been performed. VOCs with a high Henry's Law Constant would enter into the air and distribute over a large area. Whereas VOCs with a low Henry's Law Constant may persist in the sample for longer. The freeze-thawing of samples up to three times had minimal effect, but a slight increase in VOC numbers and intensities were seen indicating a re-activation of metabolism post thawing or a re-release of VOCs.

In terms of long-term storage, this work has shown that equine faecal VOCs are preserved better at -80°C than at -20°C when stored up to 12 months. It was identified that an increase in compounds, particularly esters, were characteristic at -20°C storage. For this thesis samples were either collected at the same time and stored in a -20°C or for the longitudinal study stored in a -80°C and if necessary transferred to a -20°C once all samples had been collected.

In order to identify the conditions responsible for causing ester formation at -20°C, snap freezing of samples in liquid nitrogen was performed prior to storing samples in a freezer. However, snap freezing had very little effect regardless of the vessel used and the freezer temperature the sample was then stored at. For cryovial storage, ester formation was still present at -20°C but to a lesser extent. The most compounds shared with a fresh sample was a sample stored in a cryovial at -80°C.

Despite the same animal being sampled for the storage investigations one of the main limitations was that sampling took place on four separate occasions over a 25-month period. It is anticipated that time of year may have influenced the chemical composition of samples (see **Chapter 3** for seasonal effect on faecal VOCs) therefore making direct comparisons between these individual storage investigations is difficult. Samples analysed fresh on three of the collection occasions shared 71% of compounds, indicating a change in the VOC profile

of samples between the time points sampled. The strengths of these storage investigations are that they have generated pilot results of previously unknown effects of storage conditions on equine faeces and have highlighted the importance of these factors in experimental design.

Work leading on from these pilot studies should include sampling multiple horses over time. Each sample should be subdivided into multiple vessel types and stored at various appropriate temperatures e.g. 4°C, -20°C or -80°C. This is a more robust experimental design allowing direct comparisons of storage methods to be made and statistical analysis to be performed as well as accounting for individual horse and time factors that may impact the VOC profiles.

In conclusion, these storage investigations have added further evidence demonstrating the importance of storage methods for the study of faecal VOCs. Experimental designs must consider and keep constant: sample containers, temperature and length of storage as well as time from collection to freezing and the number of times samples are defrosted and re-frozen to achieve optimal and reproducible results.

2.11 Summary of optimal method

As previously demonstrated in mice and human faeces, a unique preparation method was required for equine faeces. Based on the investigations described, the optimal method proposed for extracting VOCs using SPME-GCMS technology involves a mass of 1000 mg of equine faeces in a 10 ml vial. The faecal material should be mixed manually prior to analysis to create a homogenous sample of the inner and outer material and SPME extraction performed using a fibre coated with DVB-CAR-PDMS. Extraction time and temperature were not investigated as previously outlined due to the fact that a consistent time of 20 minutes at 60°C has been successfully applied in a range of species (Reade *et al.*, 2014).

Faeces should be ideally frozen at -80°C within four hours following spontaneous defaecation to represent VOCs that are most likely to be products of anaerobic microbial digestion. If this time frame is not practical it is important to freeze all samples at the same time to ensure VOC degradation is at the same stage to enable accurate comparisons. As demonstrated here, it is important to store all samples in the same type of container, preferably at -80°C. If -80°C storage is not available -20°C is suitable if samples are being collected and stored at the same time.

Chapter 3 Temporal variation of the faecal VOC profile and mycobiome of grazing horses

3.1 Introduction

Stability of the equine hindgut microorganisms is very important for maintaining equine health (Daly *et al.*, 2012). In terms of stability of faecal bacteria over time in healthy horses, it has been demonstrated that a concentrate diet results in a less stable faecal microbial community than a forage diet (Willing *et al.*, 2009). The faecal bacterial populations of horses and ponies on uniform diets have shown strong temporal stability (Blackmore *et al.*, 2013; Dougal *et al.*, 2017). The sampling periods carried out by Blackmore *et al*., and Dougal *et al*., were of 10 weeks and 6 weeks respectively and are currently the longest temporal studies performed. Volatile organic compounds (VOCs) emitted from equine faeces may be indicative of bacterial and fungal activity of the hindgut (Burnham *et al.*, 2018). The use of VOC analysis may provide valuable insights into the activity of the hindgut microbiota as a cost-effective alternative. Current knowledge of the temporal stability of the VOC profile of the horse is lacking, with a small number of studies which have employed pre- and post-sampling after a dietary change (Proudman *et al.*, 2015; Snalune *et al.*, 2019; Waring *et al.*, 2019).

Colic is a general term for signs of abdominal pain (discussed in **Chapter 1**) and can range from specific gastrointestinal disorders that are mild and self-limiting to severe and life-threatening forms. The prevalence of colic in the horse population has been recorded as 4% and in the late 1990s the estimated annual cost of colic in the US was $115.3 million (Traub-Dargatz *et al.*, 2001). Epidemiological studies have identified many risk factors for colic including: increased hours of stabling, turnout on new pasture, being a broodmare, parasite burden, recent transportation and changes in exercise program and season (Hillyer *et al.*, 2002; Archer & Proudman, 2006). A seasonal pattern to colic has been described and the increased likelihood of colic in general in the spring and autumn coincides with either times of management change or periods when horses are more likely to be intensively managed (Archer *et al.*, 2006). However, the biological events that underpin the seasonal risk for equine colic are unclear. Furthermore, certain types of equine colic including large colon displacement and large colon torsion that are often associated with excess gas production following a gut dysbiosis also demonstrate a biphasic pattern with increased likelihood in the spring and autumn (Archer *et al.*, 2006; Daly *et al.*, 2012). Given the seasonal risk factors

identified for colic, there is still very little knowledge of long-term temporal stability of the equine faecal microbiome and metabolome, but knowledge of this may help in identifying strategies for prevention.

The first aim of this chapter is to create a library of VOCs, and VOC patterns defining the healthy faecal metabolome of horses grazing on pasture over 12 months. Risk factors for colic including transportation, stabling and concentrate feed were kept to a minimum in order to study the natural fluctuations in the faecal metabolome of a group of non-intensively managed horses as accurately as possible. Temporal sampling allowed seasonal changes to be recorded and patterns of change to be assessed independent of confounding management factors. Characterisation of the bacterial microbiome of these samples was carried out previously (Salem *et al.*, 2018). Very little is known about the hindgut mycobiome of the horse and the role that these organisms may play in health and disease (Edwards, 2019). As discussed in **Chapter 1**, fungi in the equine gut can be divided into two categories: anaerobic fungi (AF) which are thought to contribute towards degradation of plant material to provide nutritional benefits to the host (Gruninger *et al.*, 2014). The other category of fungi most likely to colonise in the gut are facultative anaerobic fungi (FF). Some FF may be symbiotic and it has been suggested that they may provide benefits to the host in the form of probiotics or microbial additives in the feed of cattle (Abrão *et al.*, 2014). Other types of FF have been shown to play a role in the development of inflammatory disease in humans (Richard & Sokol, 2019). In horses, very little is known about the function of FF in the gut (Doxey *et al.*, 1990). Furthermore, there are no published data available on the temporal stability of equine gut fungi. The second aim of this chapter is to investigate the temporal faecal mycobiome of the horse. The method used was a metagenomics (18S rRNA and Internal transcribed spacer region 1) analysis on a subset of samples to study temporal changes in the fungal populations of the faeces.

3.2 Methods

3.2.1 Animals

Faecal samples were collected immediately post defecation from a group of seven grazing horses and ponies every four weeks over a period of 12 months from the 14th April 2014 to the 31st March 2015 (Salem *et al.*, 2018). Exact dates of sampling are shown in Table 3.1. Sampling began at 9am each day to ensure consistency with time of day and collection. Average temperature and average rainfall data were collected from a weather station in Ness

Botanical Gardens (source: https://sci.ncas.ac.uk/ness/). However, for the first two sampling dates the station was not functioning and instead the next nearest station was used (Ellesmere Port, source: https://www.wunderground.com/weather/gb/ellesmere-port). The animals were maintained on the same pasture and were not transported or stabled during the study period. The animals had *ad lib* access to grass and water, but were provided with haylage at times of the year when grass was in short supply. Horse G7 received a supplement of Course Mix (concentrate feed) during April and early May. Dietary information at each time of collection is shown in Table 3.1. The horses had been administered an anthelmintic (moxidectin and praziquantel) on 16th May 2014. FECs were carried out at the start of sampling (14th April 2014) and the 20th January 2015, results are detailed in Table 3.2. Horse G6 was excluded from the study because it was euthanised at the end of May 2014 following a traumatic limb injury.

Time point	Exact sampling date	Diet
April	Mon 14 Apr 14	Grass and haylage. Horse G7 received diet of Course mix
May 1	Mon 12 May 14	Grass and haylage. Horse G7 received diet of Course mix
May 2	Mon 26 May 14	Grass and haylage
June	Tue 24 Jun 14	Grass (haylage removed end of May)
July	Tue 22 Jul 14	Grass
Aug	Tue 19 Aug 14	Grass
Sept	Tue 16 Sep 14	Grass
Oct	Tue 14 Oct 14	Grass
Nov	Tue 11 Nov 14	Grass
Dec	Tue 09 Dec 14	Grass
Jan	Tue 06 Jan 15	Grass and haylage. Haylage introduced end of December
Feb	Tue 03 Feb 15	Grass and haylage
Mar 1	Tue 03 Mar 15	Grass and haylage
Mar 2	Tue 31 Mar 15	Grass and haylage

Table 3.1 Exact sampling dates and dietary information of horses sampled for faecal VOCs. Samples were collected once every four weeks over a 12-month period.

Horse ID	Age (years)	Breed	Sex	Faecal egg count 1	Faecal egg count 2
G1	4	Warmblood	Gelding	325	1125
G2	10	Warmblood	Mare	0	0
G3	6	Thoroughbred	Mare	825	800
G4	4	Welsh pony	Mare	475	250
G5	5	Welsh pony	Mare	2000	200
G6	3	Warmblood	Gelding	300	-
G7	3	Warmblood	Filly	575	725

Table 3.2 Demographic information and parasite status of horses sampled over a 12-month period for faecal VOCs.

3.2.2 Sample preparation and VOC profiling

Samples were transported to the laboratory within 2–6 hours of collection and were stored at −80^{o}C. Once all samples had been collected and DNA extracted for microbiome and mycobiome study they were transported to a second laboratory (on dry ice) where they were stored at -20^{o}C until VOC analysis. Samples were prepared for VOC analysis using the optimal method for extracting VOCs from horse faeces, as detailed in **Chapter 2**. Prepared samples then underwent HS-SPME-GCMS analysis (refer to **Chapter 2**). Three technical replicates were analysed for each sample. The running order of samples was computer generated, at random, to prevent any run-time bias.

3.2.3 Reference solution for VOC profiling

A reference solution was analysed at the same time once per week for quality assurance. Pyridine (10 µl), 2-pentanone (10 µl), benzaldehyde (10 µl) and indole (10 mg) were added to 500 ml of distilled water and dissolved. This was then stored in the fridge, with a shelf-life of three weeks. The same reference solution was used for weeks 1, 2, 3 and 4. A new solution was prepared for weeks 5 and 6. Aliquots of 100 µl were added to 10 ml vials and analysed using HS-SPME-GCMS and run in triplicate.

3.2.4 DNA extraction

DNA was extracted from faecal samples (stored at -80°C) for each horse from the following time points: May 1, July, Oct and Jan. DNA was not available for horse G7 for the July time point. DNA extractions were carried out by Dr. Shebl Salem (University of Liverpool, the methods have been described previously (Salem *et al.*, 2018).

3.2.5 Amplicon specific polymerase chain reaction (PCR)

To amplify the fungal 18S rRNA gene, the primers used were ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT - GGR AAA CTC ACC AGG TCCA G and GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T - GSW CTA TCC CCA KCA CGA for forward and reverse, respectively (HPLC grade, IDT) (Liu *et al.*, 2012). To amplify the internal transcribed spacer (ITS1) region, the following primers were used: ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT - TCC TAC CCT TTG TGA ATT TG and GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T - CTG CGT TCT TCA TCG TTG CG for forward and reverse, respectively (HPLC grade, Integrated DNA Technologies) (Tuckwell *et al.*, 2005). Sections of primer sequence highlighted in red are overhang which are complementary to the barcoded set of primers used in the second round of PCR. A PCR master mix was made containing (per reaction): Q5 reaction buffer 1x, 0.02 U/µl Q5 High-Fidelity DNA polymerase, 200 µM dNTPs (NEB), 0.125 µM of each forward and reverse primers made up to 15 µl with molecular grade water making a total PCR reaction volume of 20 µl. The master mix was divided between wells, along with 10 ng (2 ng/µl) of template DNA. Each plate was loaded into a pre-heated thermocycler (Multigene™ and Multigene™ OptiMax, Labnet International). For 18S rRNA the initial denaturation was 30 seconds at 98°C followed by 20 cycles of 10 seconds at 98°C, 30 seconds at 62°C, and 20 seconds at 72°C, before finishing with a 2 minutes extension at 72°C and holding at 4°C until purification (performed within 24 hours). For ITS1 the PCR conditions were the same with the exception of the annealing temperature was 55°C instead of 62°C. The first round of PCR was carried out in duplicate to reduce PCR bias. The PCR product was then stored at 4°C until purification.

3.2.6 Purification with Axygen beads

The purification step was carried out within 24 hours of PCR completion. To each well of PCR product, 20 µl of Axygen magnetic beads (AxyPrep Mag PCR clean up Kit) were added and mixed. Following an incubation of 5 minutes the plate was placed on a magnet and the liquid was removed from each well. Ethanol (200 µl, 80%) was added, then removed 30 seconds

later and repeated. The plate was then air dried for 5 minutes before removal from the magnet. The magnetic beads were then re-suspended and mixed in water, before a 1-minute incubation on the magnet. The liquid from each sample was then collected and frozen at -20^{o}C.

3.2.7 Index polymerase chain reaction (PCR)

Index PCR was carried out in a volume of 20 µl with the same reagents and method used in section 3.2.5. The entire volume of purified PCR product from section 3.2.6 was used in the reaction. An exception of the method in 3.2.5 was the use of a set of barcoded index primers (listed in Appendix 3.1) used at 0.25 µM per sample and 15 PCR cycles were performed instead of 20. Once again, this step was carried out in duplicate. The purification protocol in section 3.2.6 was performed post PCR. Prior to the purification the products from both plates were pooled and therefore the volumes used in section 3.6.2 were doubled. The final elute was 25 µl and the concentration of DNA was recorded using a Qubit (Qubit dsDNA HS assay kit, Life Technologies).

3.2.8 Sample pooling and MiSeq Illumina next generation sequencing

Prior to sample pooling an aliquot was run on a bioanalyzer (2100 Bioanalyzer, Agilent) at the Centre for Genomic Research (CGR), University of Liverpool. DNA concentration and fragment size were then used to pool samples. The pooled sample was submitted to the CGR facility where PippinPrep and qPCR quality control steps were performed prior to sequencing on an Illumina MiSeq platform.

3.2.9 Data processing and statistical analysis of VOC data

Data processing and identification of VOCs was performed as described in **Chapter 2.** For quality control purposes a PCA was performed to check for GCMS runtime clustering (Figure 3.1a). Samples that were run during week one and two clustered away from the rest (Figure 3.1a). The MS lost connection with the computer software at the end of week two and had to be re-set, which may explain why the data prior to this is clustering away from the rest. For this reason, week one and two data were removed from the analysis. Figure 3.1b shows the data after week one and two were removed. After the removal of week one and two data, data processing was performed as described in **Chapter 2**. Table 3.3 shows the final sample set. Statistical analysis included the construction of PCA plots labelled for horse, feed type, time point, season, average environmental temperature and rainfall. To evaluate the association of VOC abundance with factors of interest (time point, feed type, season, average

environmental temperature and average rainfall) were included in a linear mixed effects (LME) model. Horse ID was set as the random variable. A heatmap of compounds identified as significantly associated with variables of interest was constructed to demonstrate the change in compound abundance over time. A PERMANOVA analysis was also performed to evaluate the contribution of the factors horse, time point and feed type to variation in the VOC profile. As FEC was only recorded on two occasions (Table 3.2) the relationship of FEC with VOC profile was investigated using the samples where FEC was performed or the next sampled analysed for VOCs after a FEC was performed (April and Feb). Thresholds for low, medium and high FECs were determined using the guidelines outlined previously (Kaplan & Nielsen, 2010; Nielsen *et al.*, 2013). The association of FEC and VOC profile was investigated using PCA and PERMANOVA analysis.

3.2.10 Data processing of 18S rRNA and ITS1 data

Illumina adapter sequences were trimmed from raw FastQ files using Cutadapt (1.2.1) (Martin, 2011). The reads were further trimmed using Sickle (1.2) (Joshi & Fass, 2011) with a minimum window quality score of 20. Reads shorter than 200 bp after trimming were removed. Reads with fewer than 1000 reads were also removed at this stage. Sequence clustering (99% similarity) was performed by SWARM 2.0 (d=3) (Mahé *et al.*, 2015) and chimeras were filtered out using UCHIME (Edgar *et al.*, 2011). The following steps were then carried out using Quantitative Insights into Microbial Ecology (QIIME) (1.9.1) (Caporaso *et al.*, 2010b). The classifier tool BLAST (Edgar, 2010) was applied together with the Silva database (18S rRNA) (Quast *et al.*, 2013) or Unite database (ITS1) (Nilsson *et al.*, 2019) to assign OTUs at a 99% threshold. The final steps performed consisted of alignment of OTUs with the database using PyNAST (Caporaso *et al.*, 2010a) and construction of a phylogenetic tree (for 18S rRNA as databases are currently lacking for ITS1) using FastTree (Price *et al.*, 2010). Dr Alessandra Frau (University of Liverpool) performed the data processing steps for the 18S rRNA and ITS1 data sets.

3.2.11 Mycobiome statistical analysis

The following statistical analysis was performed for both 18S rRNA and ITS1 datasets. For visualisation purposes, taxonomy plots were constructed for phylum, class, order and family. The data were normalised to relative abundance for these visualisation plots. Diversity indices were calculated at OTU level using the R package vegan. The alpha diversity (Richness, Shannon and Fisher indexes) was calculated for each group and differences between groups were compared using the aov() function in R. The beta diversity was plotted on non-metric

multidimensional scaling (NMDS) ordination plots (distances: Bray, Unifrac and Weighted Unifrac). Statistical differences in the beta diversity between groups was assessed by PERMANOVA using the adonis() function. Dr Umer Zeeshan Ijaz (University of Glasgow) wrote the taxonomy, diversity and PERMANOVA scripts.

Associations between OTU abundance and variables of interest (time point and feed type) were evaluated using LME modelling. The data was normalised using Total Sum Scaling (TSS) followed by Centered Log Ratio (CLR) prior to LME modelling.

3.2.12 Integration of mycobiome and metabolome data

The R package mixOmics using the Data Integration Analysis for Biomarker discovery (DIABLO) framework was used to identify correlations between OTUs and VOCs. Specifically, the N-integrative supervised analysis was performed with DIABLO (Rohart *et al.*, 2017). The model was performed twice: first to include VOC and 18S rRNA and second to include ITS1 data instead of 18S rRNA. Data integration was performed on time points where VOCs, mycobiome and microbiome data sets were available. Prior to input of data into the DIABLO model, 18S rRNA and ITS1 data were normalised by TSS followed by CLR. VOC data were normalised by methods described in **Chapter 2.** The first step involved fitting the model prior to variable selection to assess performance (based on correct assignment of groups) using the function perf(). A 10-fold cross validation of perf() was undertaken. From the cross-validation, a performance plot was constructed displaying the overall classification error rate and balanced error rate (BER). The balanced error rate is preferential when unbalanced groups are used because it calculates the average proportion of wrongly classified samples in each class, weighted by the number of samples in each class. An error rate of less than 30% is considered acceptable for the performance of PLS-DA models (Bijlsma *et al.*, 2006). Based on the performance of the model the optimal number of components was chosen using the function choice.ncomp(). Next, the tune.block.splsda function was used to choose the optimum number of variables for each type of omics (OTUs and VOCs). Using these optimal variables, the final DIABLO model was constructed. A scatterplot (Pearson's correlation) was generated for OTUs and VOCs separately to see which omics was the most important in distinguishing between time points using the function plotIndiv(). The cor.test function (for paired data) was used to determine significant pair-wise correlations identified between OTUs and VOCs. P values were adjusted using the Benjamini-Hochberg correction to correct for multiple comparisons (Benjamini & Hochberg, 1995). A plot generated by the plotDIABLO() function was used to show overall correlation between selected OTUs and

VOCs. The plotLoadings() function was used to visualise loading weights of each of the selected variables on each component for both OTUs and VOCs. For further correlation analysis a circle plot was constructed using the plotVar() function. The circle plot allows correlations to be visualised based on the angles of two vectors (González *et al.*, 2012). A sharp angle represents a positive correlation, an obtuse angle represents a negative correlation and a null correlation is observed when there is a right angle. For visualisation of specific OTU-VOC correlations a circos plot was constructed using the circosPlot() function. Dr Umer Zeeshan Ijaz (University of Glasgow) wrote the mixOmics R script.

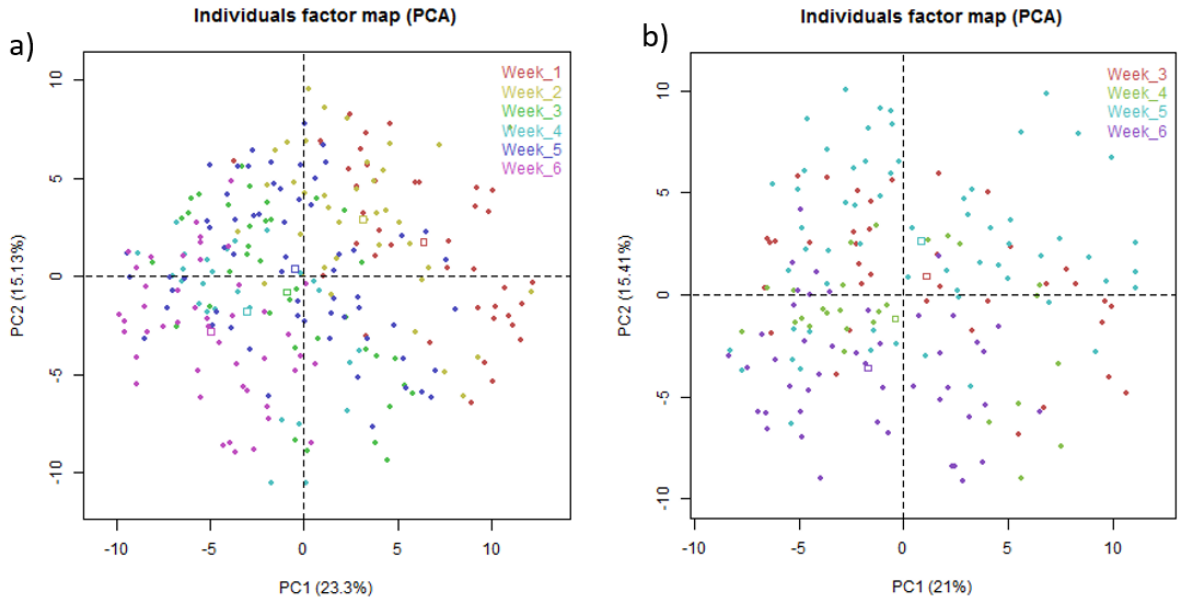


Figure 3.1 PCAs of the VOC profiles of all faecal samples before a) and after b) removal of week one and week two data. Faecal samples were collected from horses grazing at pasture over 12 months.

Horse	APR	MAY1	MAY2	JUNE	JULY	AUG	SEPT	OCT	NOV	DEC	JAN	FEB	MAR1	MAR2
G1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Im	✓	✓
G3	✓	✓	✓	✓	Im	✓	✓	✓	✓	✓	✓	✓	✓	✓
G4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
G7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 3.3 Faecal samples collected from horses. Samples marked with a (✓) show at least one technical replicate was available following the removal of data run on the GCMS the first and second week. Data was imputed (based on an average of samples within that time point) for samples marked with (Im). Mycobiome analysis was performed on highlighted samples.

3.3 Results

3.3.1 VOC metabolome results

3.3.1.1 Reference solution

PCAs for reference solution are shown in Figure 3.2 and PCA loading scores are listed in Table 3.4. In a PCA where reference solution data was included for all weeks, week 2 in particular was an outlier from the rest (Figure 3.2a). In a PCA with week 1 and 2 data removed, clearer separation for reference solution batch was evident (Figure 3.2b). The CoV was calculated to assess overall compound stability, interestingly when week 2 data were removed the CoV improved for all compounds apart from 2-pentanone (Table 3.4). According to PCA loading scores 2-pentanone was responsible the main separation along PC1 indicating differences between the two batches of reference solution.

	All weeks of reference solution			**Weeks 1 and 2 reference solution excluded**		
Compound	CoV (%)	PC1 scores	PC2 scores	CoV (%)	PC1 scores	PC2 scores
2-pentanone	13.84	0.50	0.46	15.42	-0.43	-0.90
Pyridine	3.93	-0.59	0.35	3.38	0.52	-0.24
Benzaldehyde	3.40	-0.37	-0.65	2.84	0.53	-0.23
Indole	9.39	-0.52	0.50	7.68	0.51	-0.27

Table 3.4 Coefficient of variation and PCA loading scores (PC1 and PC2) for reference solution

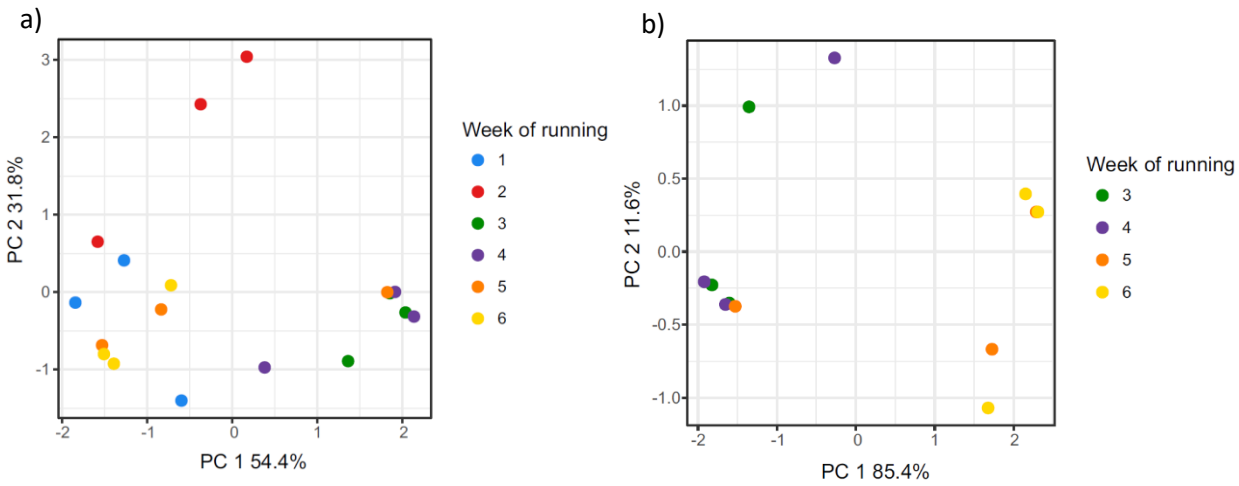
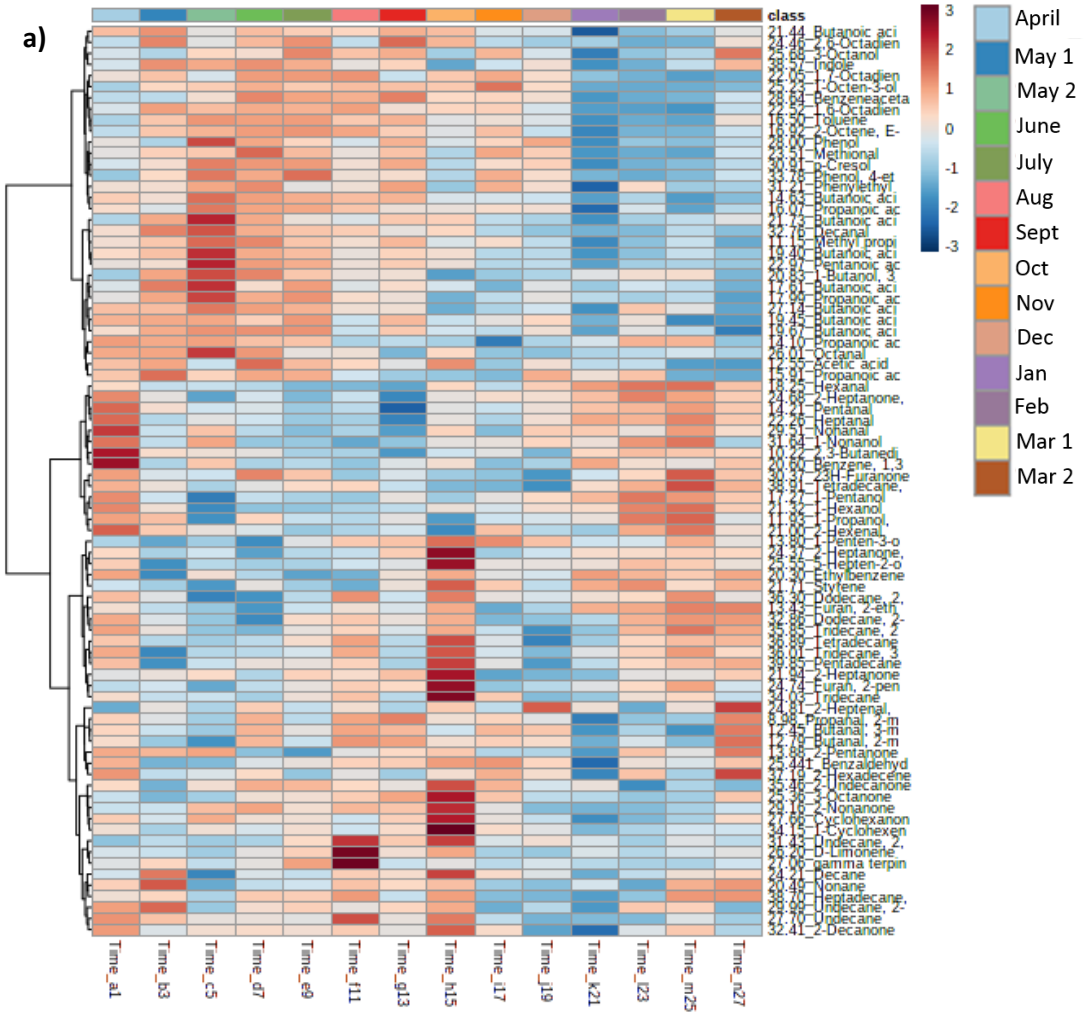


Figure 3.2 Reference solution PCAs before a) and after b) removal of week 1 and 1 week 2 data.

3.3.1.2 VOCs identified and core compounds

A total of 134 VOCs was identified across all 14 time points and 6 horses. Core compounds (present in at least one sample per time point) represented 63% of the VOC profile. A heatmap of core compounds is shown in Figure 3.3a and a full list of compounds and their frequency in samples is in Appendix 3.2. A stacked plot of five time points selected to represent change in chemical composition of compounds between seasons is shown in Figure 3.3b. The stacked plot was weighted by compound presence within samples.



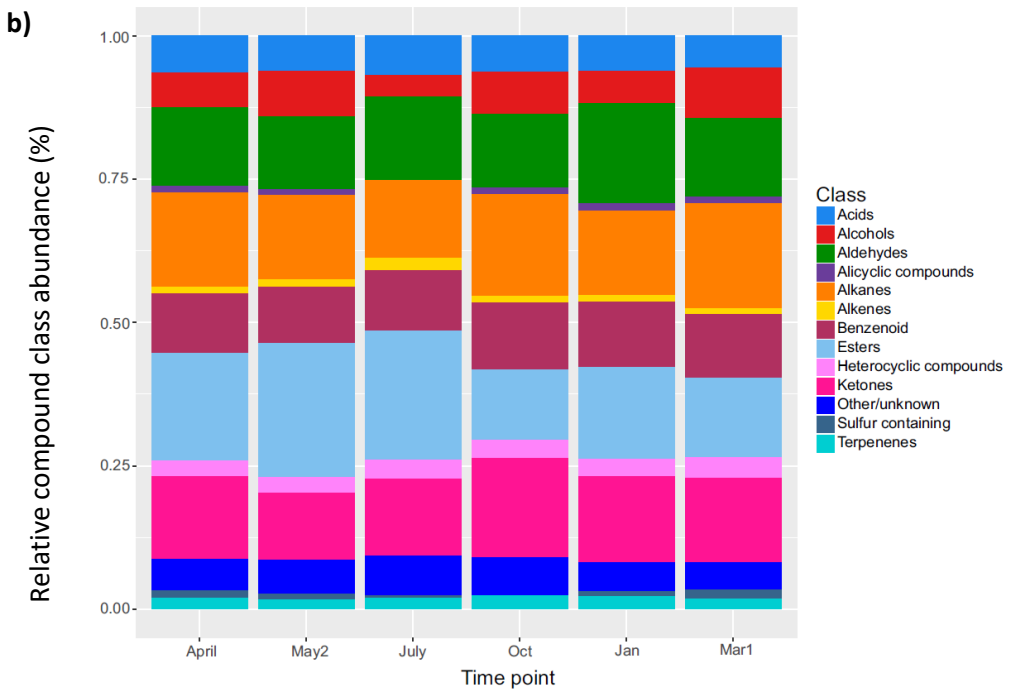


Figure 3.3 a) Heatmap of core compounds and b) chemical composition of the faecal VOC library created in this study. For b) chemical classes were weighted by number of times VOCs belonging to each chemical class were present in samples at the time point displayed on the x-axis. Some key VOCs of various chemical classes were plotted in Appendix 3.3 to show their variability. Faecal samples were collected every four weeks from horses grazing at pasture over 12 months.

3.3.1.3 PCA plots

PCA plots were constructed to investigate clustering in the VOC profile for the variables: horse, feed type, time point, season, average temperature and average rainfall (Figures 3.4a-f). Clustering was not evident for horse or average rainfall (Figure 3.4a and Figure 3.4f, respectively). Some clustering for season was observed in a PCA; with summer samples towards the left and winter samples towards the right of the plot (Figure 3.4d). Separation in the VOC profile for feed type was fairly clear, with the exception of May1 and May2 samples which were clustering with the grass samples, despite the horses also having access to haylage during this time (Figure 3.4b). In Figure 3.4c a linear shift in VOC profile with time point was evident, with winter and spring clustering to the right of the plot (Figure 3.4c). In terms of average temperature, a linear pattern of samples with change in temperature was apparent across the PCA plot (Figure 3.4e). A table of the ten VOCs responsible for the most variation in the positive and negative directions of the PC1 is shown in Table 3.5.

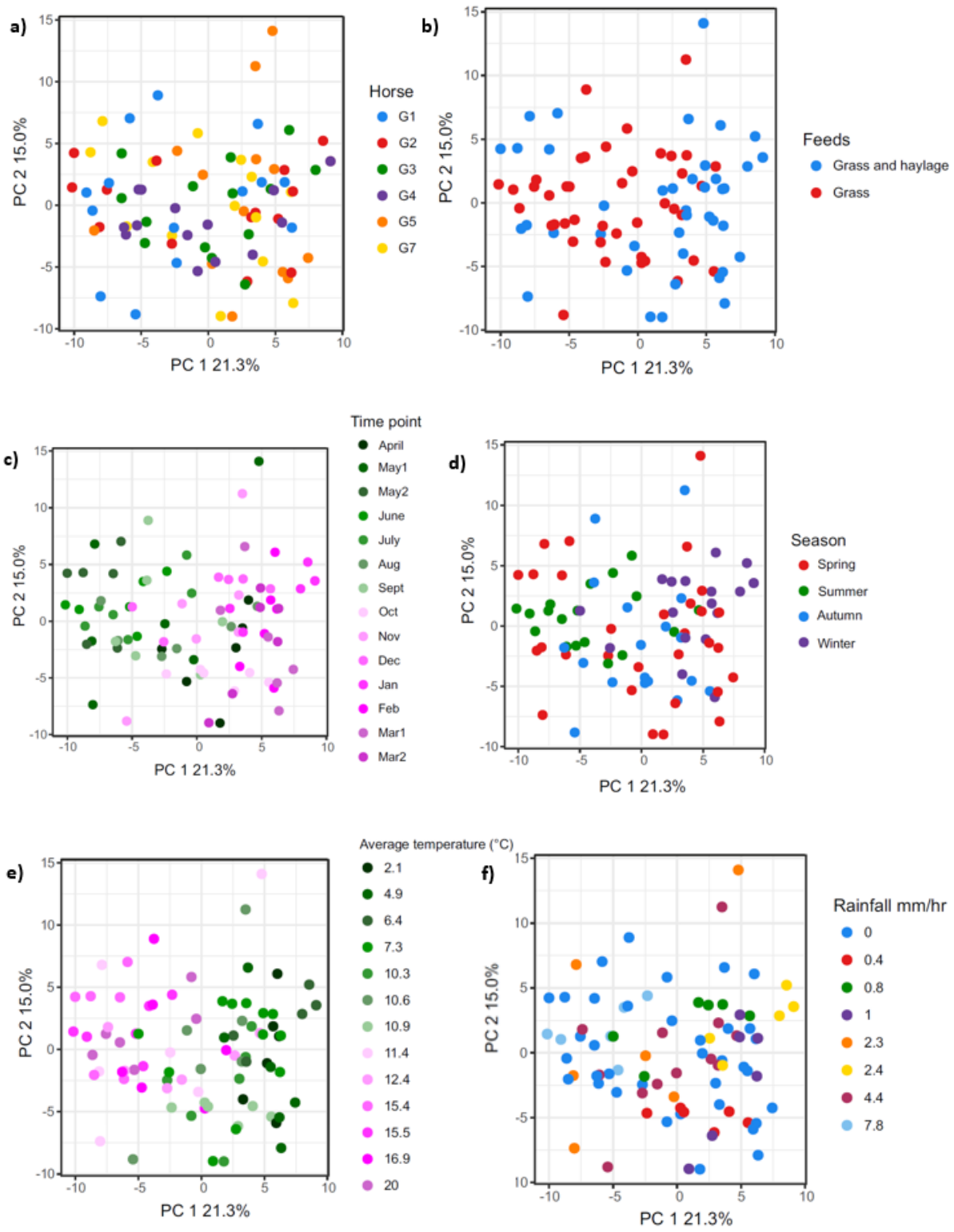


Figure 3.4 PCA plots of faecal VOC profiles labelled for the variables: horse a), feed type b), time point c), season d) average temperature e) and average rainfall f). Faecal samples were collected from horses grazing at pasture over 12 months.

PC1 positive direction		PC1 negative direction	
VOC	**score**	**VOC**	**score**
2-Heptanone, 5-methyl-	0.06	Pentanoic acid, butyl ester	-0.16
Nonane, 2-methyl-	0.07	Butanoic acid, 1-methylethyl ester	-0.16
Oxime-, methoxy-phenyl-	0.07	Pentanoic acid, methyl ester	-0.16
Hexanal	0.08	Butanoic acid, methyl ester	-0.16
1-Pentanol	0.08	p-Cresol	-0.16
1-Butanol, 2-methyl-	0.08	Propanoic acid, 2-methyl-, propyl ester	-0.15
1-Propanone, 1-cyclopropyl-	0.09	n-Propyl acetate	-0.15
Octane	0.11	Propanoic acid, propyl ester	-0.15
1-Hexanol	0.12	1-Butanol, 2-methyl-, acetate	-0.15
1-Heptanol	0.13	Pentanoic acid, 2-methylbutyl ester	-0.15

Table 3.5 The ten highest loading scores VOCs of VOCs in the positive and negative directions of PC1 from Figure 3.4a-f. Faecal samples were collected from horses grazing at pasture over 12 months.

3.3.1.4 The association of VOC abundance and factors of interest

Tables 3.6 lists 10 VOCs that are significantly associated with variables of interest pre- and post-correction for multiple comparisons. After correction for multiple comparisons, 12, 16, 29 and 2 VOCs were significantly associated with time point, feed type, average temperature and average rainfall, respectively. One VOC (heptadecane, 2,6,10,14-tetramethyl-) was significantly associated with season. A heatmap to show change in abundance of significant VOCs over time is shown in Figure 3.5. Of the variables recorded in this study, a PERMANOVA analysis revealed time point as the variable accountable for the most variation in VOC profile (Table 3.11).

VOC associated with time point	p-value	Adjusted p-value
19.45_Butanoic acid, 2-methyl-, ethyl ester	<0.001	<0.001
32.76_Decanal	<0.001	<0.001
22.05_1,7-Octadiene, 2,7-dimethyl-	<0.001	0.015
22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	0.001	0.015
37.80_2-Undecenal	0.001	0.015
10.60_Ethyl acetate	0.001	0.015
29.00_Propanoic acid, hexyl ester	0.001	0.015
29.81_Decane, 2,9-dimethyl-	0.002	0.027
25.33_1-Butanol, 3-methyl-, propanoate	0.002	0.028
26.01_Octanal	0.003	0.036
35.70_Undecanal	0.004	0.043
14.63_Butanoic acid, methyl ester	0.004	0.043
30.32_Pentanoic acid, 3-methylbutyl ester	0.004	0.045
19.67_Butanoic acid, 3-methyl-, ethyl ester	0.006	0.054
25.23_1-Octen-3-ol	0.006	0.054
23.58_Butanoic acid, 2-methylpropyl ester	0.008	0.060
23.84_Benzene, propyl-	0.008	0.060
28.27_Pentanoic acid, butyl ester	0.008	0.060
25.14_1-Hepten-3-one	0.009	0.060
27.14_Butanoic acid, 2-methylbutyl ester	0.013	0.083
31.64_1-Nonanol	0.015	0.095
25.68_3-Octanol	0.018	0.108
7.19_Propanal	0.019	0.108
29.99_Undecane, 2-methyl-	0.020	0.111
10.80_2-Butanol	0.024	0.127
15.91_Propanoic acid, 2-methyl-, ethyl ester	0.025	0.127
17.99_Propanoic acid, propyl ester	0.026	0.127
12.55_Acetic acid	0.032	0.149
14.10_Propanoic acid, ethyl ester	0.033	0.149
16.66_Octane	0.035	0.154
21.44_Butanoic acid, 3-methyl-	0.039	0.165

Table 3.6 VOCs associated with the factor time point. Faecal samples were collected from horses grazing at pasture over 12 months. The retention time of each VOC is recorded before each VOC name.

VOC associated with feed type	p-value	Adjusted p-value
13.94_1-Penten-3-ol	<0.001	<0.001
14.17_3-Pentanone	<0.001	<0.001
25.23_1-Octen-3-ol	<0.001	<0.001
25.441_Benzaldehyde	<0.001	<0.001
34.15_1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	<0.001	0.008
14.30_n-Propyl acetate	0.001	0.010
22.05_1,7-Octadiene, 2,7-dimethyl-	0.001	0.010
22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	0.001	0.012
25.36_3-Octanone	0.001	0.012
25.14_1-Hepten-3-one	0.001	0.016
9.89_Furan, 2-methyl-	0.001	0.016
36.30_Dodecane, 2,6,10-trimethyl-	0.002	0.024
32.41_2-Decanone	0.002	0.024
25.33_1-Butanol, 3-methyl-, propanoate	0.003	0.024
31.43_Undecane, 2,6-dimethyl-	0.003	0.024
13.80_1-Penten-3-one	0.004	0.031
25.55_5-Hepten-2-one, 6-methyl-	0.007	0.051
21.44_Butanoic acid, 3-methyl-	0.007	0.054
27.66_Cyclohexanone, 2,2,6-trimethyl-	0.008	0.054
26.20_D-Limonene	0.009	0.059
28.64_Benzeneacetaldehyde	0.011	0.070
21.69_Propanoic acid, butyl ester	0.012	0.071
16.92_2-Octene, (E)-	0.016	0.087
23.58_Butanoic acid, 2-methylpropyl ester	0.017	0.087
23.84_Benzene, propyl-	0.017	0.087
24.21_Decane	0.018	0.087
24.74_Furan, 2-pentyl-	0.018	0.087
27.06_gamma terpinene	0.018	0.087
29.81_Decane, 2,9-dimethyl-	0.019	0.089
12.79_Butanal, 2-methyl-	0.023	0.100
16.93_1-Propanone, 1-cyclopropyl-	0.025	0.105
14.10_Propanoic acid, ethyl ester	0.026	0.105
10.50_2-Butanone	0.026	0.105
34.03_Tridecane	0.027	0.105
20.83_1-Butanol, 3-methyl-, acetate	0.028	0.105
8.98_Propanal, 2-methyl-	0.034	0.125
38.70_Heptadecane, 2,6,10,14-tetramethyl-	0.035	0.125
17.99_Propanoic acid, propyl ester	0.038	0.131
20.60_Benzene, 1,3-dimethyl-	0.039	0.131
23.27_2-Octene, 2,6-dimethyl-	0.043	0.142

Table 3.7 VOCs associated with the factor feed type. Faecal samples were collected from horses grazing at pasture over 12 months. The retention time of each VOC is recorded before each VOC name.

VOCs associated with season	p-value	Adjusted p-value
38.70_Heptadecane, 2,6,10,14-tetramethyl-	<0.001	0.013
32.41_2-Decanone	0.002	0.090
20.49_Nonane	0.002	0.090
13.94_1-Penten-3-ol	0.003	0.090
25.441_Benzaldehyde	0.005	0.106
23.27_2-Octene, 2,6-dimethyl-	0.005	0.106
25.68_3-Octanol	0.006	0.106
35.85_Tridecane, 2-methyl-	0.007	0.108
22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	0.007	0.109
21.44_Butanoic acid, 3-methyl-	0.010	0.128
16.92_2-Octene, (E)-	0.013	0.160
24.21_Decane	0.018	0.197
27.66_Cyclohexanone, 2,2,6-trimethyl-	0.021	0.199
32.76_Decanal	0.022	0.199
26.53_Dimethyl sulfone	0.022	0.199
21.32_1-Hexanol	0.029	0.199
38.57_Indole	0.030	0.199
13.88_2-Pentanone	0.030	0.199
8.98_Propanal, 2-methyl-	0.032	0.199
10.50_2-Butanone	0.033	0.199
32.86_Dodecane, 2-methyl-	0.034	0.199
18.01_2-Hexanone	0.034	0.199
21.00_2-Hexenal	0.034	0.199
36.30_Dodecane, 2,6,10-trimethyl-	0.040	0.223
36.89_Tetradecane	0.044	0.225
19.45_Butanoic acid, 2-methyl-, ethyl ester	0.044	0.225

Table 3.8 VOCs associated with season. Faecal samples were collected from horses grazing at pasture over 12 months. The retention time of each VOC is recorded before each VOC name.

VOCs associated with average temperature	p-value	Adjusted p-value
14.30_n-Propyl acetate	<0.001	<0.001
16.50_Toluene	<0.001	<0.001
16.66_Octane	<0.001	<0.001
19.13_Butanoic acid, 1-methylethyl ester	<0.001	<0.001
21.32_1-Hexanol	<0.001	<0.001
24.68_2-Heptanone, 5-methyl-	<0.001	<0.001
24.90_1-Heptanol	<0.001	<0.001
21.69_Propanoic acid, butyl ester	<0.001	0.002
14.21_Pentanal	<0.001	0.003
35.73_1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	<0.001	0.003
35.70_Undecanal	<0.001	0.003
36.30_Dodecane, 2,6,10-trimethyl-	<0.001	0.003
22.73_Pentanoic acid, 1-methylethyl ester	0.001	0.005
30.44_Pentanoic acid, 2-methylbutyl ester	0.001	0.009
24.06_Oxime-, methoxy-phenyl-	0.001	0.011
33.78_Phenol, 4-ethyl-	0.002	0.014
24.46_2,6-Octadiene, 2,6-dimethyl-	0.002	0.018
31.64_1-Nonanol	0.003	0.021
18.25_Hexanal	0.003	0.021
18.64_Pentanoic acid, methyl ester	0.003	0.021
17.99_Propanoic acid, propyl ester	0.004	0.025
24.93_Pentanoic acid, propyl ester	0.004	0.025
13.94_1-Penten-3-ol	0.005	0.026
19.65_Propanoic acid, 2-methyl-, propyl ester	0.006	0.032
20.94_1-Butanol, 2-methyl-, acetate	0.006	0.032
16.23_1-Butanol, 2-methyl-	0.008	0.039
25.441_Benzaldehyde	0.008	0.039
25.55_5-Hepten-2-one, 6-methyl-	0.009	0.042
28.27_Pentanoic acid, butyl ester	0.009	0.042
14.17_3-Pentanone	0.011	0.049
29.51_Nonanal	0.013	0.056
17.27_1-Pentanol	0.016	0.068
10.50_2-Butanone	0.022	0.089
16.92_2-Octene, (E)-	0.024	0.095
30.91_p-Cresol	0.027	0.104
21.00_2-Hexenal	0.029	0.105
23.22_Butanoic acid, 3-methyl-, propyl ester	0.030	0.106
22.26_Heptanal	0.030	0.106
26.53_Dimethyl sulfone	0.033	0.111
32.41_2-Decanone	0.038	0.121
23.58_Butanoic acid, 2-methylpropyl ester	0.039	0.121
23.84_Benzene, propyl-	0.039	0.121
23.27_2-Octene, 2,6-dimethyl-	0.039	0.121
22.97_Pentanoic acid	0.042	0.126
x11.93_1-Propanol, 2-methyl-	0.043	0.127

Table 3.9 VOCs associated with average environmental temperature. Faecal samples were collected from horses grazing at pasture over 12 months. The retention time of each VOC is recorded before each VOC name.

VOCs associated with average rainfall (mm/hour)	p-value	Adjusted p-value
24.37_2-Heptanone, 6-methyl-	<0.001	0.040
13.94_1-Penten-3-ol	0.001	0.040
34.15_1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.002	0.064
21.32_1-Hexanol	0.002	0.064
25.55_5-Hepten-2-one, 6-methyl-	0.002	0.064
15.94_Disulfide, dimethyl	0.005	0.088
21.94_2-Heptanone	0.006	0.088
31.64_1-Nonanol	0.006	0.088
9.89_Furan, 2-methyl-	0.007	0.088
32.86_Dodecane, 2-methyl-	0.007	0.088
16.66_Octane	0.007	0.088
29.81_Decane, 2,9-dimethyl-	0.008	0.091
27.66_Cyclohexanone, 2,2,6-trimethyl-	0.014	0.140
17.27_1-Pentanol	0.016	0.156
10.50_2-Butanone	0.018	0.160
13.80_1-Penten-3-one	0.020	0.170
36.30_Dodecane, 2,6,10-trimethyl-	0.030	0.233
25.441_Benzaldehyde	0.032	0.233
21.69_Propanoic acid, butyl ester	0.036	0.251
19.13_Butanoic acid, 1-methylethyl ester	0.041	0.272
24.74_Furan, 2-pentyl-	0.043	0.273

Table 3.10 VOCs associated with average rainfall. Faecal samples were collected from horses grazing at pasture over 12 months. The retention time of each VOC is recorded before each VOC name.

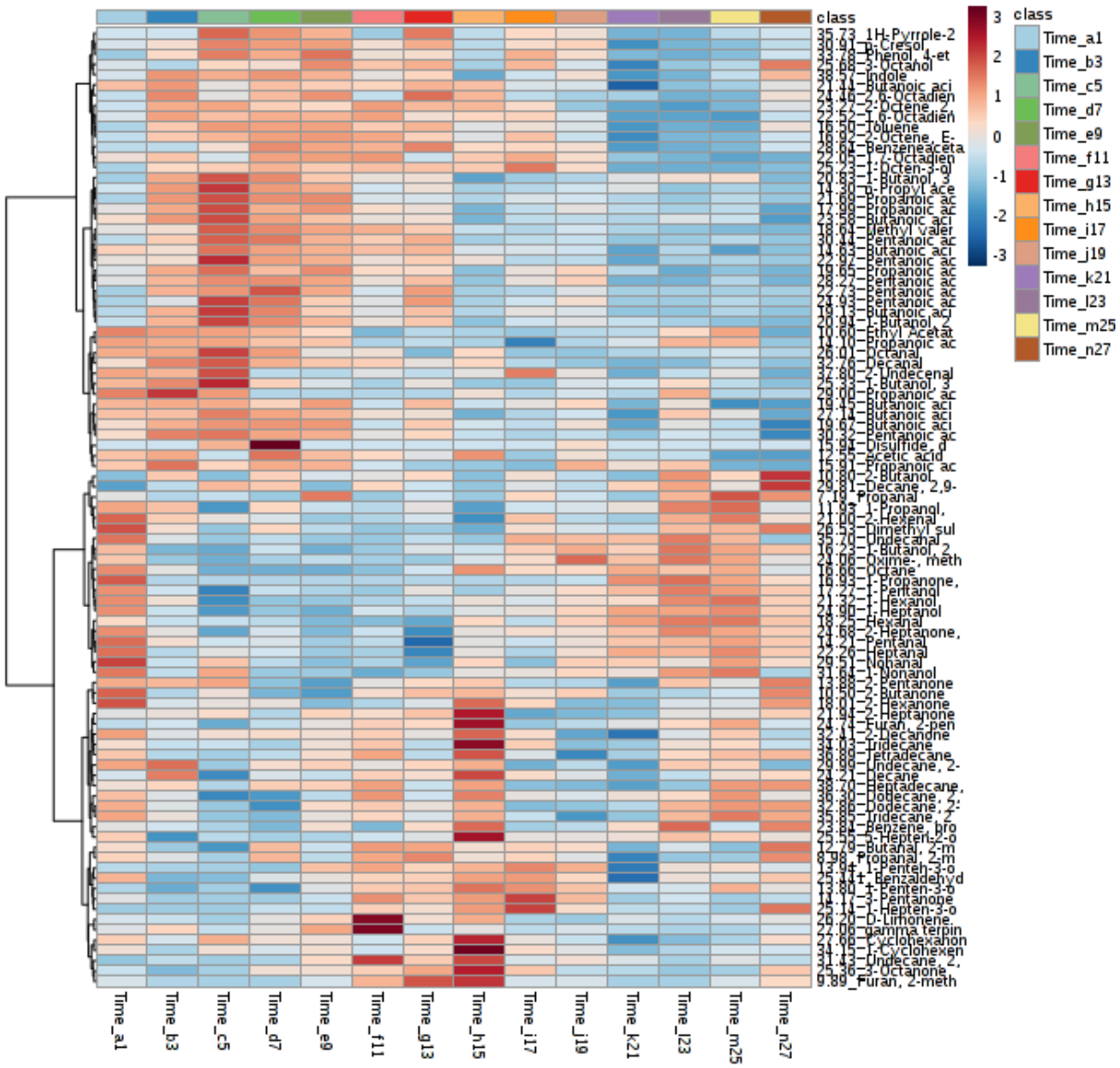


Figure 3.5 Heatmap of VOCs identified as significant in LME. Some VOC names are truncated and can be identified by retention time in tables above. Faecal samples were collected once every four weeks from horses grazing at pasture over 12 months.

Factor	% variation	p-value
Time point	14%	0.001
Feed type	7%	0.001
Horse	10%	0.001

Table 3.11 Results of a PERMANOVA analysis of the faecal VOC profiles of horses grazing at pasture over 12 months.

3.3.1.5 Relationship of FEC and VOC profile

As FEC was only recorded on two occasions (Table 3.2), the relationship of FEC and VOC profile was investigated for the time points when both of these samples were taken. PCAs of April and Feb were performed and grouping for FEC thresholds (low, medium and high) investigated (Figure 3.6). Clusters were most apparent for time point rather than FEC burden or horse (Figure 3.6). A PERMANOVA analysis mirrored these results: FEC (R^2 = 0.06, p = 0.53), time point (R^2 = 0.19, p = 0.02) and horse (R^2 = 0.12, p = 0.19).

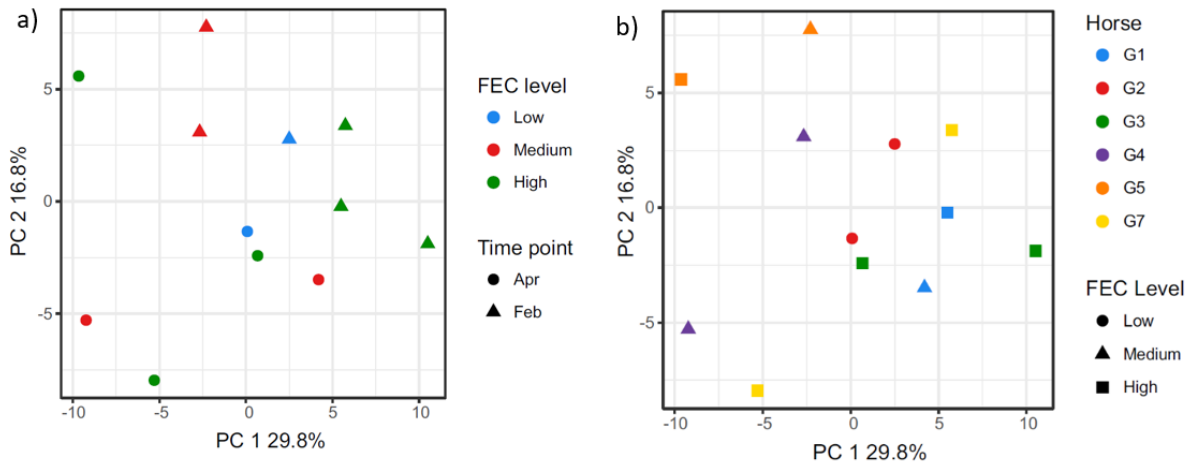


Figure 3.6 PCAs labelled for faecal egg count (FEC) level with time point a) and horse with FEC b). Horses were grazing at pasture over 12 months and faeces were analysed for VOCs and FEC on the time points shown in these PCAs. FEC levels were defined in eggs per gram (e.p.g) as low (0-199 e.p.g), medium (200-499 e.p.g) and high >500 e.p.g)

3.3.2 Mycobiome results: 18S rRNA gene

From time points May1 (T1), July (T2), Oct (T3) and Jan (T4) 22 samples were successfully sequenced with the 18S rRNA primer set. Two samples were excluded from the analysis. DNA was not available for horse G7 for July and for horse G2, the July sample failed the filtering steps outlined in section 3.2.10 and was not taken forward to statistical analysis. 1,456,866 sequences were obtained post filtering.

The sequences clustered to a total number of 289 OTUs. A minimum of 41,546, and maximum of 225,624, sequences were observed per sample (mean = 66,221). Prior to statistical analysis 79 OTUs not belonging to kingdom fungi were removed.

3.3.2.1 Taxonomic summaries: 18S rRNA gene

Visual taxonomic summaries for phylum, class, order and family for all samples are shown in Figures 3.7a-d. In total 8 phyla, 19 classes, 56 orders and 85 families of fungi were identified. The most dominant phyla were Ascomycota (43.5%), Neocallimastigomycota (42.2%) and Basidiomycota (11.8%). In terms of change in abundance of the two major phyla between time points, Ascomycota accounted for 31.6%, 47.4%, 75.1% and 24.6% for T1, T2, T3 and T4, respectively. Neocallimastigomycota was responsible for 56.6%, 41.4%, 1.7% and 65.2% of total abundance for T1, T2, T3 and T4, respectively. Genus level was not included in taxonomic summaries or in the following statistics because 40% of OTUs could not be identified to genus level.

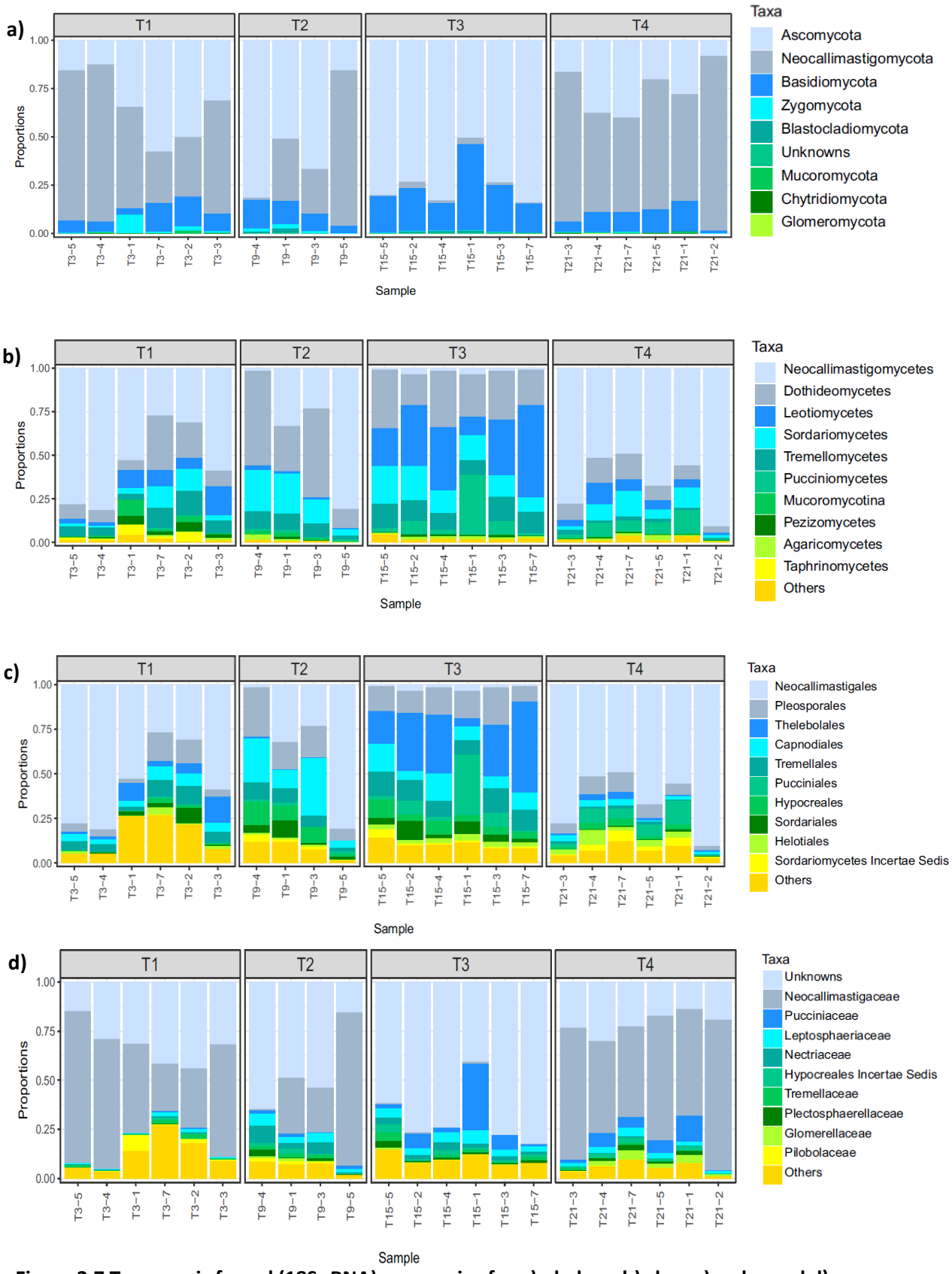


Figure 3.7 Taxonomic fungal (18S rRNA) summaries for a) phylum, b) class c) order and d) family of faeces collected from six horses over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). Samples for two horses (G2 and G7) were missing from T2.

3.3.2.2 Diversity indices: 18S rRNA gene

Box plots of the alpha diversity indices (Richness, Shannon and Fisher alpha) are shown in Figure 3.8a. The alpha diversity of T3 was significantly greater than T1, T2 and T4 for both Fisher and Richness (Figure 3.8a) For Shannon T3 was significantly greater in diversity than T4 only (Figure 3.8a). Time point had a significant effect on beta diversity with the following PERMANOVA results, R^2 = 0.52, p=0.001 (bray), R^2 = 0.33, p=0.001 (Unifrac), R^2 = 0.60, p=0.001 (weighted Unifrac). An NMDS plot of the beta diversity (bray distance) is shown in Figure 3.8b showing distinct clustering of samples for time point.

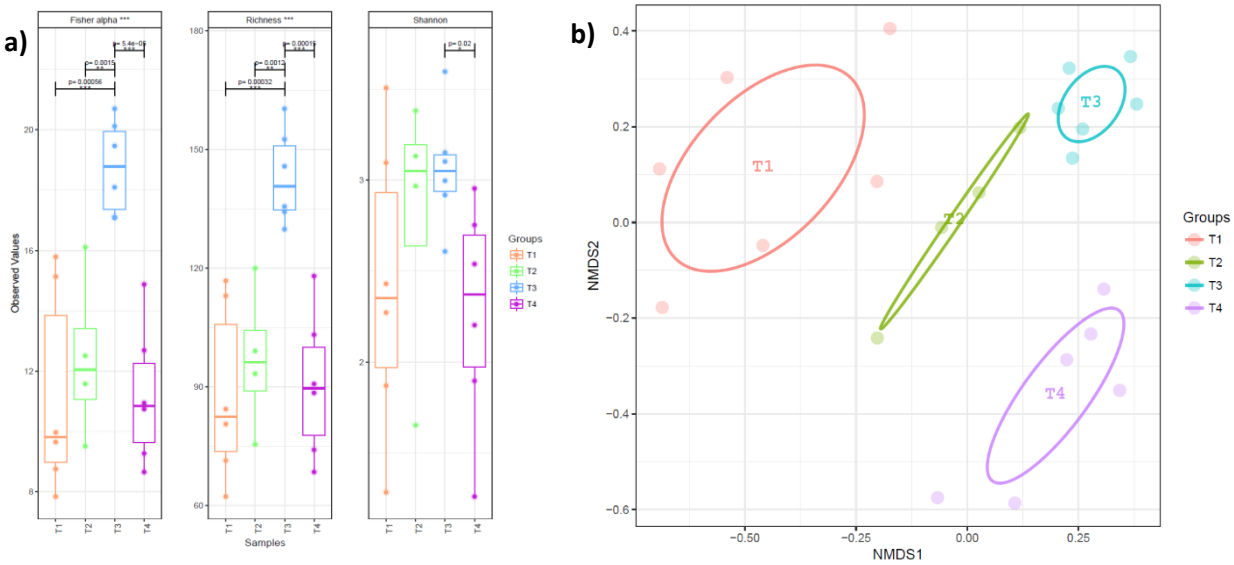


Figure 3.8 a) The alpha diversity and b) the beta diversity (bray distance) of the mycobiome (18S rRNA) of faecal samples collected from six horses over four time points which represented spring (T1), summer (T2), autumn (T3) and winter (T4). Significant differences in alpha diversity are shown on the plot (a). Beta diversity was significantly different between time points with an R^2 of 0.52 (p = 0.001, PERMANOVA).

3.3.2.3 Linear mixed effects model results: 18S rRNA gene

LME results for phylum, class, order and family for are shown in Table 3.12 for time point and Table 3.13 for feed type. Boxplots of all taxa associated with the variables time and feed type are shown in Appendix 3.4. Forty-five OTUs were significantly associated with time point and 25 were significantly associated with feed type. The 45 OTUs were included in a heatmap shown in Figure 3.9.

Time point	p-value	FDR
Phylum		
Zygomycota	<0.01	0.01
Class		
Tremellomycetes	<0.01	<0.01
Pucciniomycetes	<0.01	<0.01
Eurotiomycetes	<0.01	0.01
Taphrinomycetes	<0.01	0.01
Pezizomycetes	0.01	0.03
Mucoromycotina	0.01	0.04
Sordariomycetes.Incertae.Sedis	<0.01	<0.01
Order		
Chaetothyriales	<0.01	<0.01
Tremellales	<0.01	<0.01
Pucciniales	<0.01	<0.01
Microbotryales	<0.01	<0.01
X..Glomerellales	<0.01	<0.01
Helotiales	<0.01	0.01
Mucorales	<0.01	0.02
Taphrinales	<0.01	0.02
Sordariales	0.01	0.04
Capnodiales	0.01	0.04
Family		
Chaetothyriaceae	<0.01	<0.01
Plectosphaerellaceae	<0.01	<0.01
Glomerellaceae	<0.01	<0.01
Pucciniaceae	<0.01	<0.01
Erysiphaceae	<0.01	<0.01
Lyophyllaceae	<0.01	0.01
Ustilentylomataceae	<0.01	0.01
Debaryomycetaceae	<0.01	0.01
Hypocreales.Incertae.Sedis	<0.01	0.01
Pilobolaceae	<0.01	0.03
Protomycetaceae	<0.01	0.03
Leptosphaeriaceae	0.01	0.04

Table 3.12 LME modelling results of fungal taxa identified by the 18S rRNA gene that were associated with the variable time point. Horses were sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4).

Feed type	p-value	FDR
Phylum		
Ascomycota	<0.01	<0.01
Basidiomycota	<0.01	<0.01
Neocallimastigomycota	<0.01	<0.01
Class		
Tremellomycetes	<0.01	0.01
Neocallimastigomycetes	<0.01	0.01
Dothideomycetes	<0.01	0.02
Cystobasidiomycetes	<0.01	0.02
Pezizomycetes	<0.01	0.02
Order		
Microbotryales	<0.01	0.01
Neocallimastigales	<0.01	0.01
Hypocreales	<0.01	0.03
Family		
Neocallimastigaceae	<0.01	0.03
Nectriaceae	<0.01	0.03
Coniothyriaceae	<0.01	0.03
Plectosphaerellaceae	<0.01	0.03

Table 3.13 LME modelling results of fungal taxa identified by the 18S rRNA gene that were associated with the variable feed type. Horses were sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4).

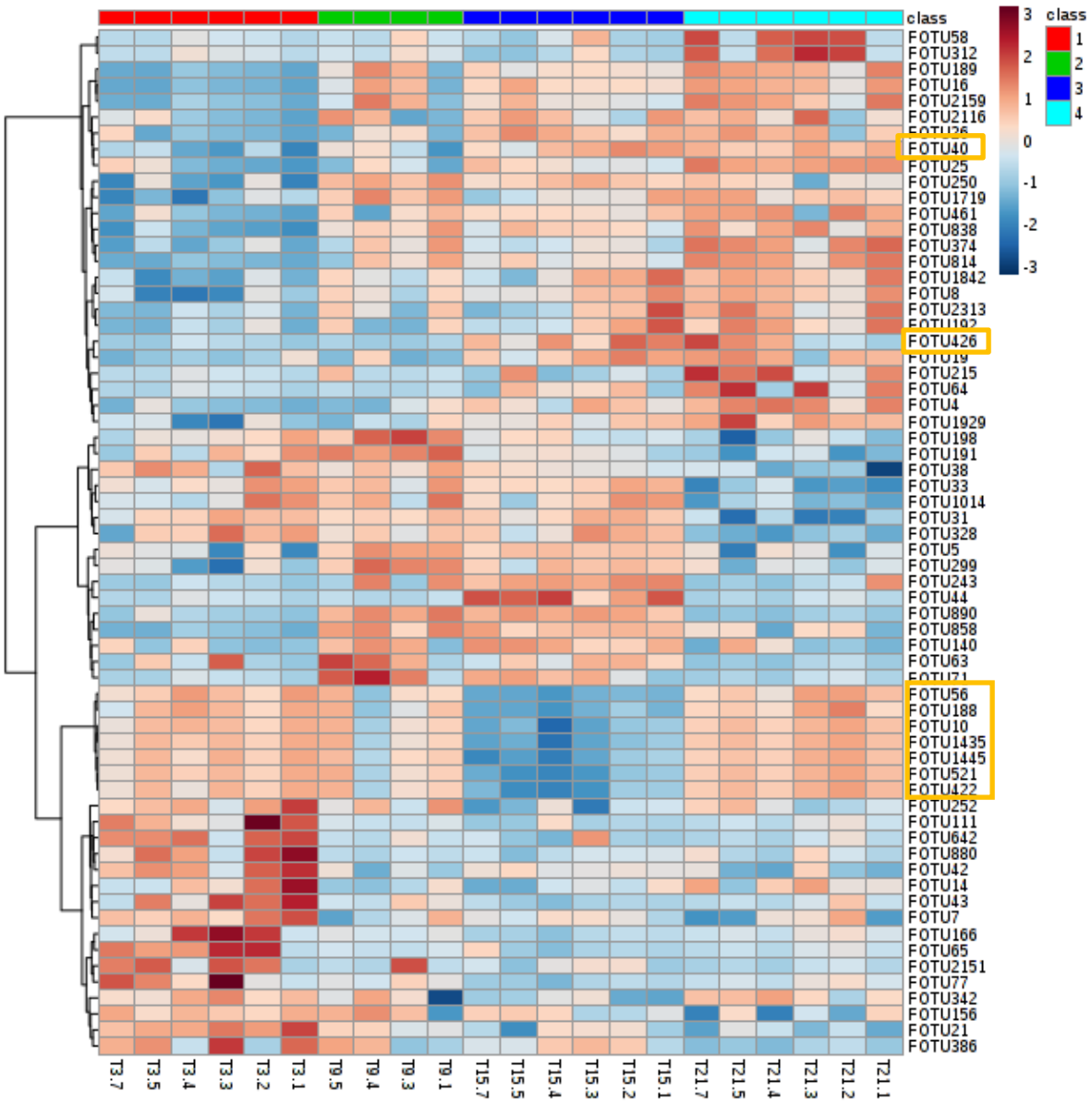


Figure 3.9 A heatmap of OTUs identified (18S rRNA) as associated with the variables 'time point' and 'feed type' in LME modelling. Anaerobic fungi OTUs are marked by a box. Horses were sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4).

3.3.3 Mycobiome results: ITS1 region of rRNA gene

The same time points were sequenced for the ITS primer set, as for the 18S rRNA, with a total of 22 samples. As with the 18S rRNA primer set, DNA was not available for horse G7 from July. One sample for horse G7 (Oct) failed the filtering steps outlined in section 3.2.10 and was not taken forward for statistical analysis. There were 1,248,285 number post filtering. There were 46 OTUs in the sequence clusters. A minimum of 1,788, and a maximum of 111,544, sequences were observed per sample (mean 56,740). Two OTUs which did not have a database match were removed prior to statistical analysis.

3.3.3.1 Taxonomy: ITS1 region of rRNA gene

Visual taxonomic summaries for phylum, class, order and family and genus for all samples are shown in Figures 3.10 and 3.11. Two phyla were identified as Neocallimastigomycota and Ascomycota. At genus level the majority of OTUs were assigned to anaerobic fungi (98.4%). The genera were unknown Neocallimastigales (36.5%), *Neocalimastix* (27.6%), *Caecomyces* (27.1%) and *Piromyces* (7.2%).

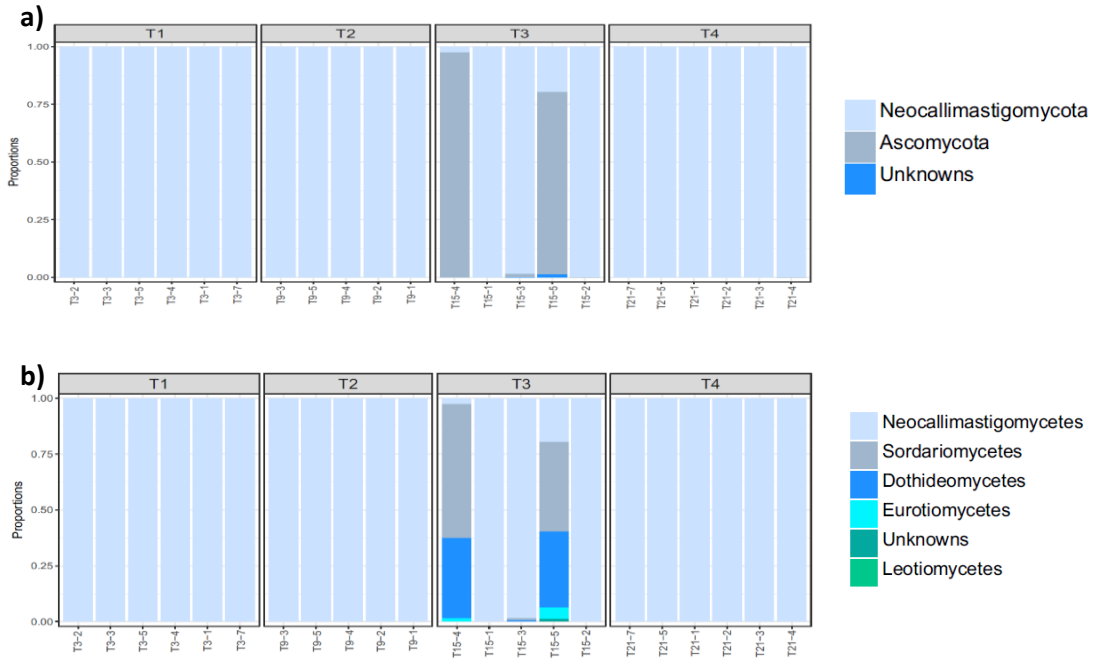


Figure 3.10 Taxonomic fungal (ITS1) summaries for a) phylum and b) class of faeces collected from six horses over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). Samples from horse G7 were missing from T2 and T3.

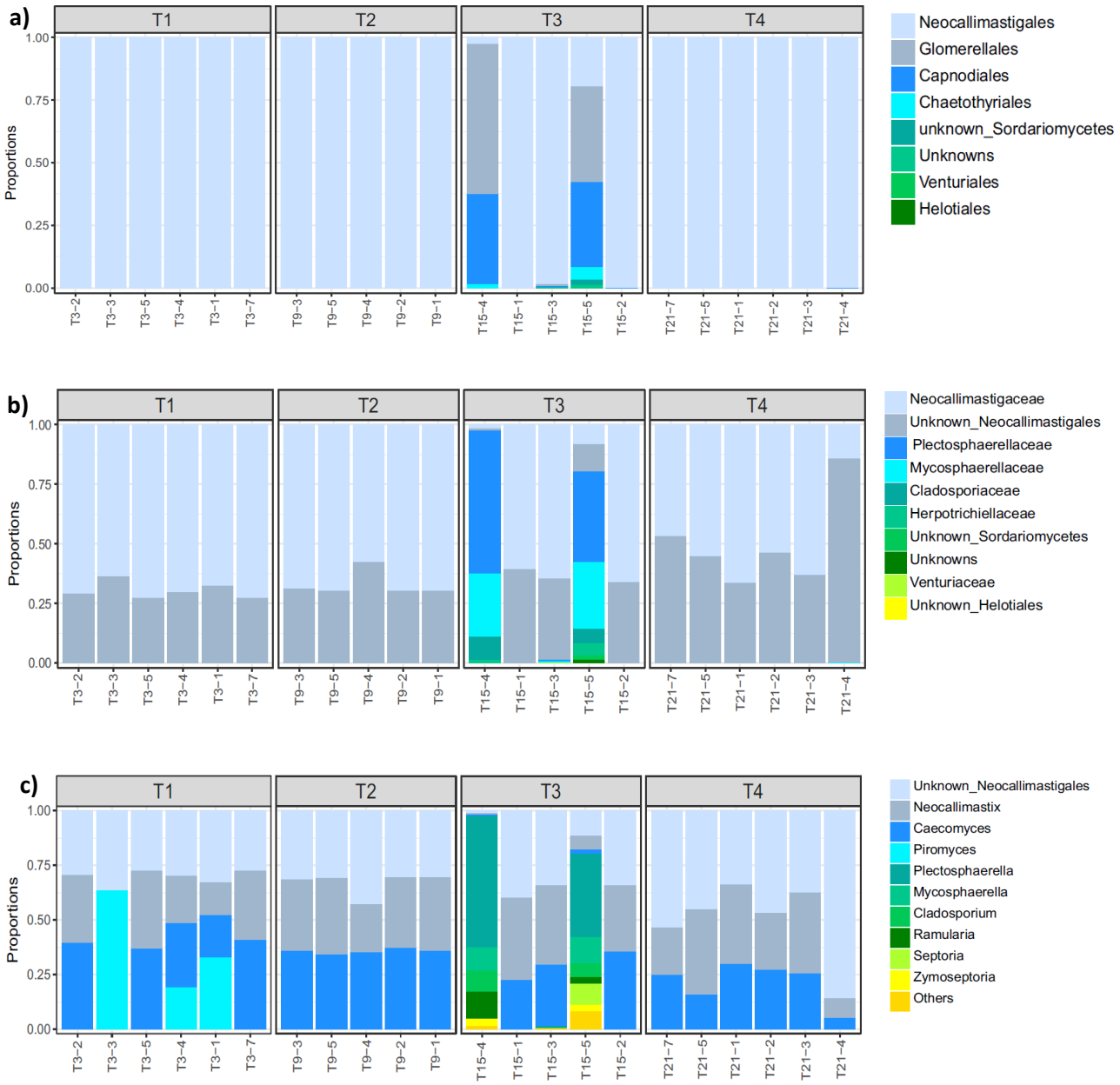


Figure 3.11 Taxonomic fungal (ITS1) summaries for a) order, b) family and c) genus of faeces collected from six horses over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). Samples from horse G7 were missing from T2 and T3.

3.3.3.2 Diversity indices: ITS1 region of rRNA gene

Box plots of the alpha diversity indices (Richness, Shannon and Fisher alpha) are shown in Figure 3.12a. The alpha diversity of T3 was significantly greater than T1, T2 and T4 for Fisher ($p<0.03$). For Richness, T3 was significantly more diverse than T4 ($p=0.04$). There were no significant differences in beta diversity when the Shannon index was used. Time point had a significant effect on beta diversity using the bray distance with an R^2 of 0.46 ($p=0.001$, PERMANOVA). An NMDS plot of the beta diversity (bray distance) is shown in Figure 3.10b showing distinct clustering of samples for time point.

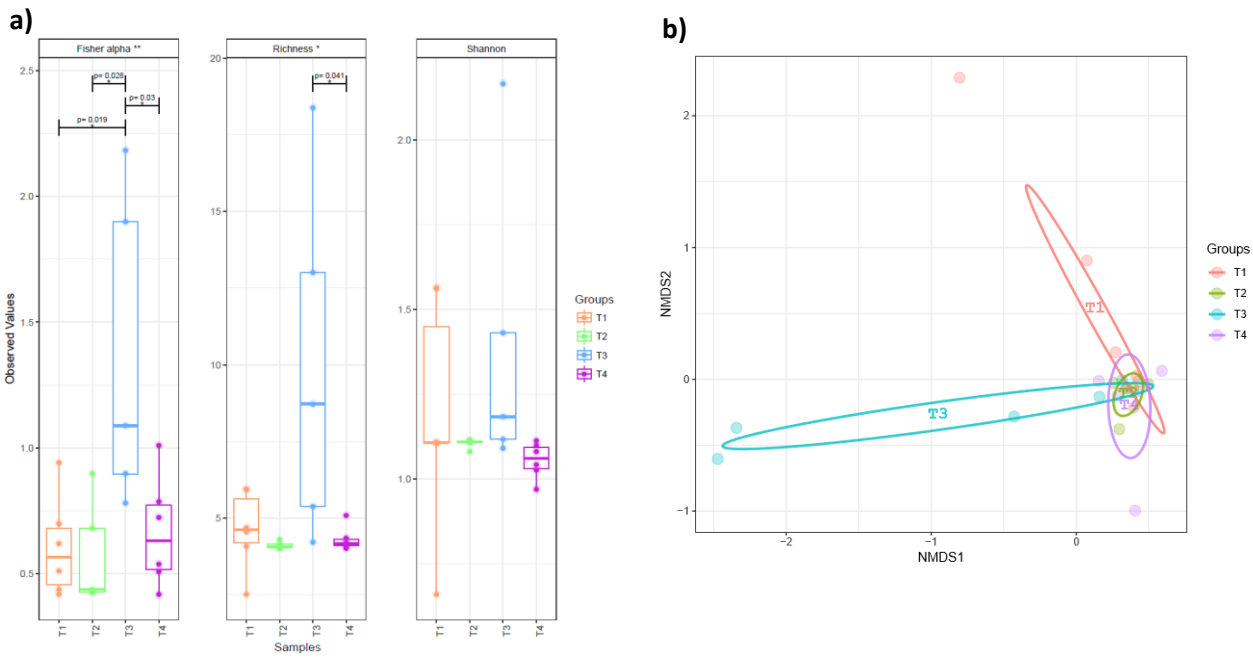


Figure 3.12 a) The alpha diversity and b) the beta diversity (bray distance) of the mycobiome (ITS1 region of rRNA gene) of faecal samples collected from six horses over four time points which represented spring (T1), summer (T2), autumn (T3) and winter (T4). Significant differences in alpha diversity are shown in plot a). Beta diversity was significantly different between time points with an R^2 of 0.46 ($p = 0.001$, PERMANOVA).

3.3.3.2 Linear mixed effects model results: ITS1 region of rRNA gene

Results of the LME model are shown in Table 3.14. After correction for multiple comparisons at phylum, class, order and family levels there were no taxa significantly associated with the variables time point or feed type. At OTU level, one OTU (OTU 20, genus *Piromyces*) was significantly associated with time point (p=0.04, FDR) (Figure 3.13c). Boxplots of the phyla Ascomycota and Neocallimastigomycota were plotted to enable comparisons to be made with the patterns observed in these phyla with the 18S rRNA primer set (Figure 3.13a and b).

Phylum	**Time point p-value**	**FDR**	**Feed type p-value**	**FDR**
Neocallimastigomycota	0.18	0.18	0.05	0.05
Ascomycota	0.06	0.11	0.05	0.05
Class				
Sordariomycetes	0.09	0.21	0.04	0.08
Leotiomycetes	0.12	0.21	0.03	0.08
Dothideomycetes	0.12	0.21	0.06	0.08
Neocallimastigomycetes	0.28	0.35	0.06	0.08
Eurotiomycetes	0.64	0.64	0.60	0.60
Order				
Glomerellales	0.08	0.34	0.04	0.11
Capnodiales	0.11	0.34	0.06	0.11
Helotiales	0.15	0.34	0.03	0.11
Neocallimastigales	0.31	0.54	0.06	0.11
unknown_Sordariomycetes	0.40	0.55	0.39	0.47
Venturiales	0.54	0.63	0.47	0.47
Chaetothyriales	0.84	0.84	0.46	0.47
Family				
Cladosporiaceae	0.91	0.99	0.40	0.48
Venturiaceae	0.67	0.99	0.66	0.66
Neocallimastigaceae	0.20	0.40	0.10	0.20
Plectosphaerellaceae	0.08	0.31	0.03	0.17
Mycosphaerellaceae	0.10	0.31	0.06	0.17
Herpotrichiellaceae	0.99	0.99	0.31	0.46

Table 3.14 LME modelling results of fungal taxa identified by the ITS1 region of rRNA gene. Horses were sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4).

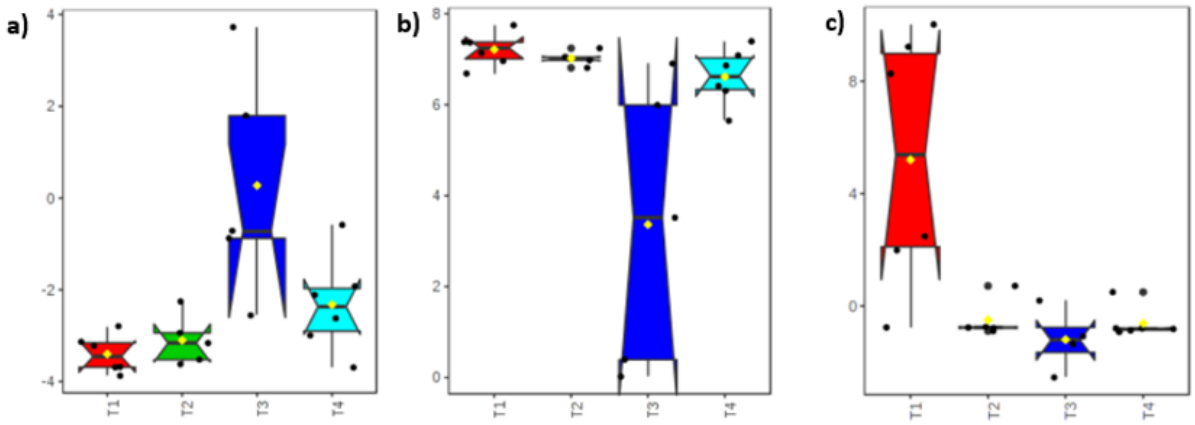


Figure 3.13 Boxplots of Ascomycota (a), Neocallimastigomycota (b) and OTU 020 (genus *Piromyces*) (c) identified by the ITS1 region of rRNA gene. Horses were sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4).

3.3.4 Integration of VOC and fungal OTU data

3.3.4.1 VOCs and 18S rRNA

A performance plot was constructed from cross-validation of the model displaying the overall classification error rate and BER (Figure 3.14). Figure 3.15 shows component 1 and component 2 of a Pearson's correlation plot to show how well fungal OTUs and VOCs can discriminate between time points (Figure 3.15a and b). Stronger clustering for time point was observed for fungi rather than VOCs (Figure 3.15c and d). The loadings for each component (demonstrating which variables are important in each group) are shown in Figure 3.16 a (fungi) and b (VOCs). In Figure 3.17a is a circle plot which shows correlations between OTUs and VOCs. Clusters of points indicate strong correlations between variables. Based on the circle plot a circos plot was constructed to visualise specific correlations between OTUs and VOCs as well as the groups they were more or less abundant in (Figure 3.17b). For a clearer representation of the circos plot, a heatmap was also produced in Figure 3.17c to show positive and negative correlations (0.3 and above) between variables. Labels of OTUs included in the loadings for the Pearson's correlation plot (Figure 3.15a) and for the circos plot (Figure 3.17b) are in Appendix 3.5.

Correlations between specific VOCs and OTUs were also generated from the Pearson's correlation plot. A total of 219 significant correlations (0.6 and above) were identified between OTUs and VOCs. Of the OTUs identified as significantly associated with variables of interest in single omics analysis, 37 out of 66 correlated with VOCs when the omics were integrated. Nine OTUs belonged to AF and were significantly correlated with VOCs, represented 98.7% of AF sequences identified by 18S rRNA. For seven AF OTUs a number of common correlations with VOCs were observed (Table 3.15). Two AF OTUs (OTU426 and OTU40) did not correlate to the same VOCs as the other AF OTUs. A heat map analysis performed earlier (Figure 3.9) revealed OTU426 and OTU40 did not follow the same pattern of change over time as the rest. Furthermore, a phylogenetic tree analysis revealed OTU426 and OTU40 were branched separately to the other AF OTUs (screenshot of phylogenetic tree is in Appendix 3.6). For FF OTUs correlations with VOCs were more variable between OTUs, a list of correlations of (>0.6) is shown in Table 3.16. However, one VOC (1-cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-) positively correlated with four different OTUs.

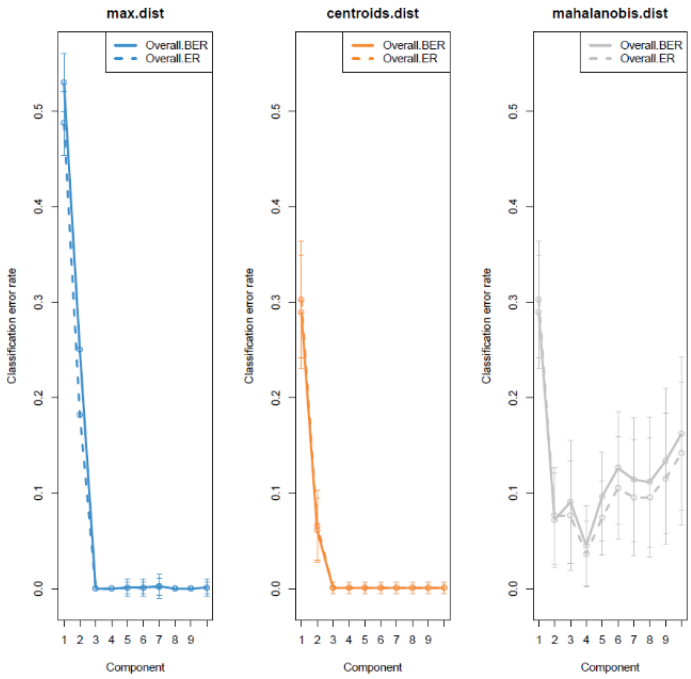


Figure 3.14 The classification error rate of a model used to combine fungi (18S rRNA) and VOC data of the faeces of horses sampled over four time points. Key: ER = error rate, BER = balanced error rate.

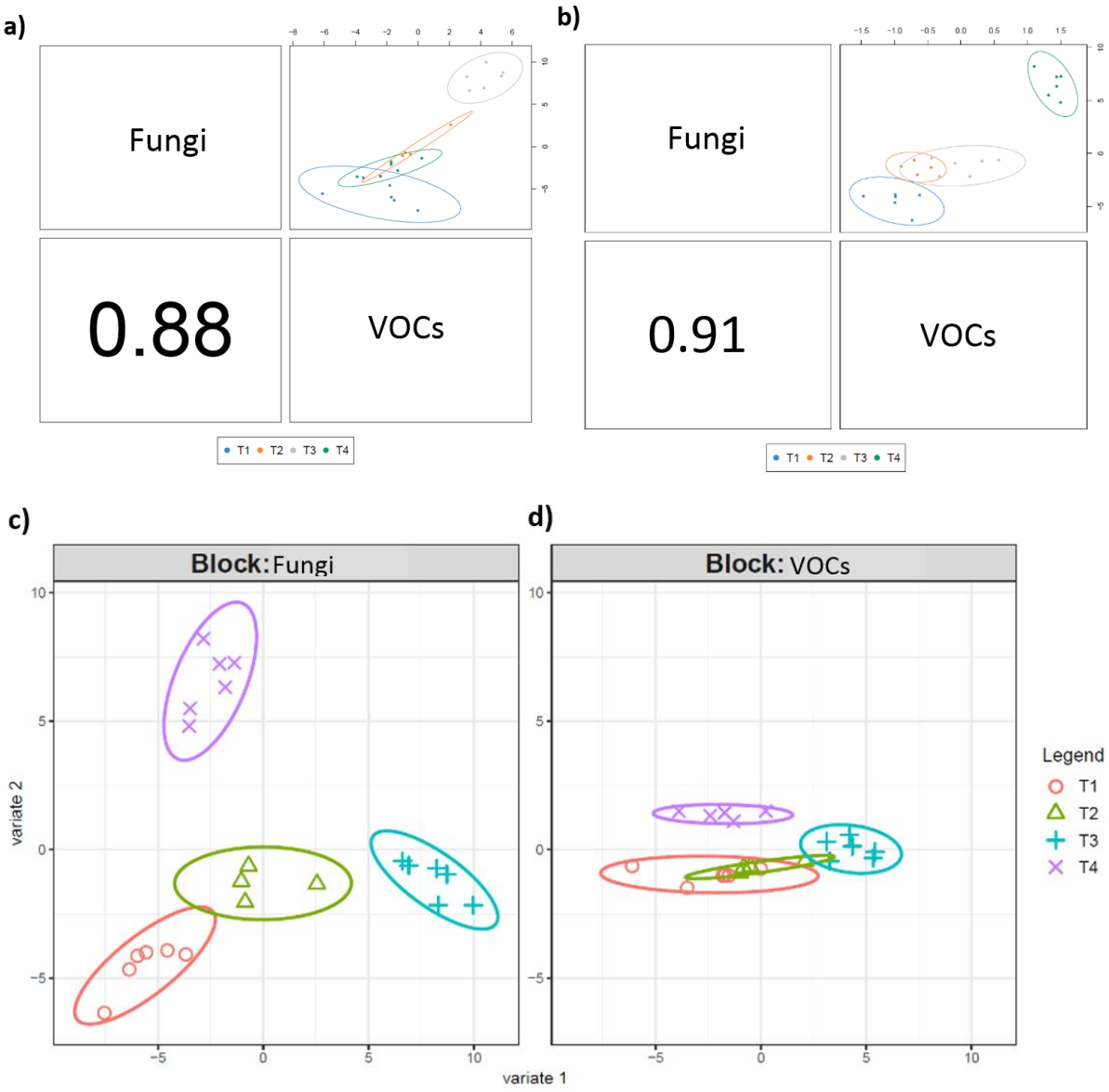


Figure 3.15 A Pearson's correlation plot of fungi (18S rRNA) and VOC data of the faeces of horses sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). Component 1 is shown in **a** and component 2 is shown in **b**. The ability of the model to separate time points by fungi alone is demonstrated by plot **c** and VOCs alone by plot **d**.

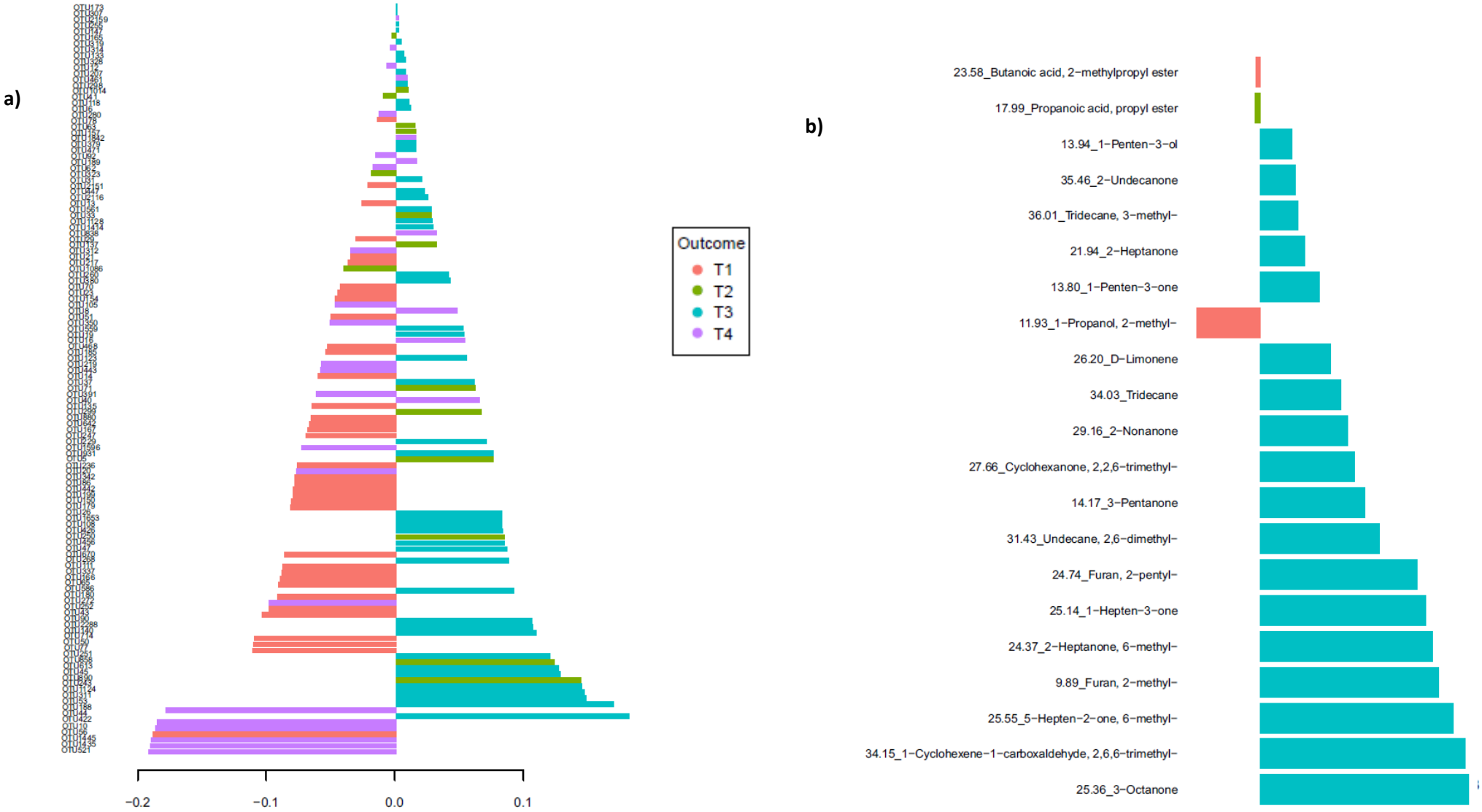
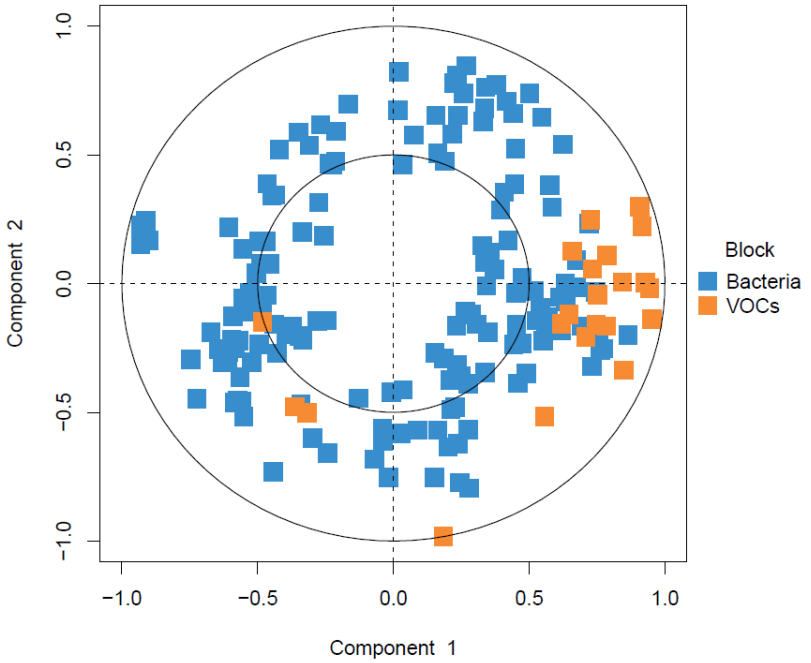
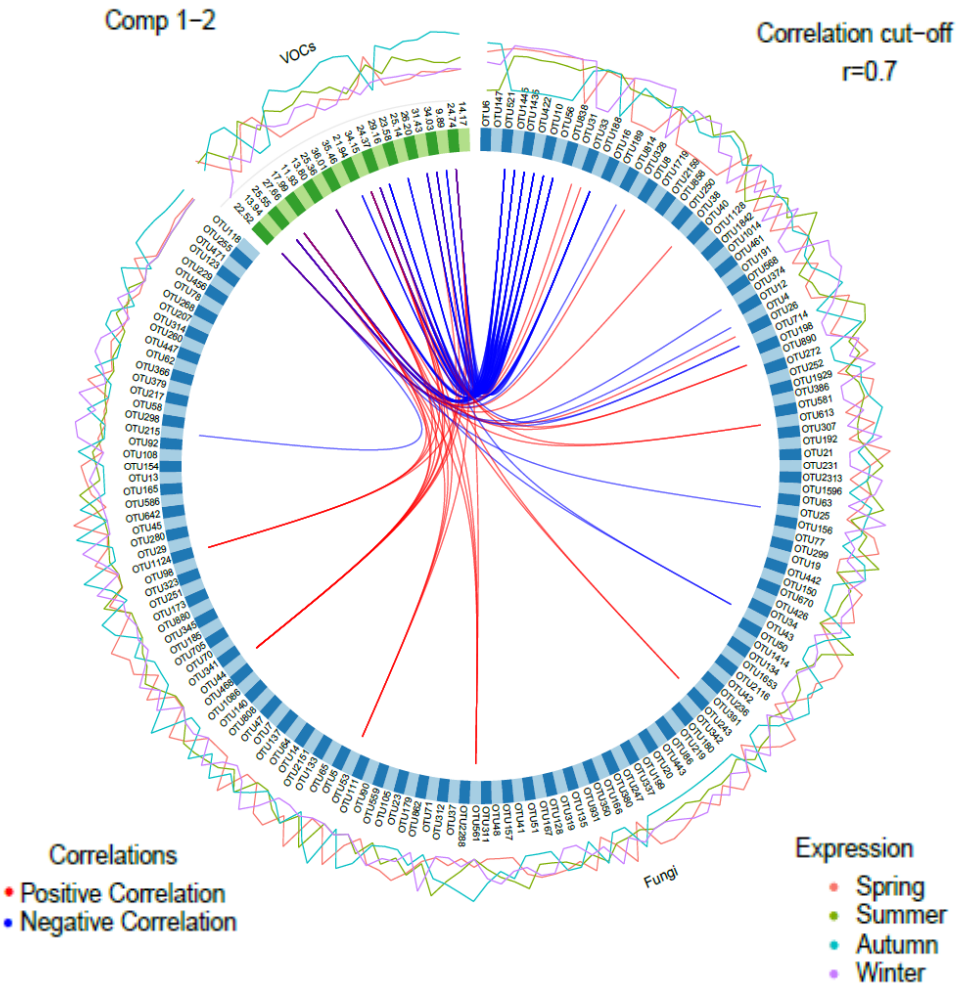


Figure 3.16 Loadings for component 1 of the Pearson's correlation plot (Figure 5.15a) of fungi and VOC data of the faeces of horses sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). OTU labels can be found in Appendix 3.5.

a)


Component 2
Component 1
1.0
0.5
0.0
−0.5
−1.0
Block
Bacteria
VOCs

b)


Comp 1−2
Correlation cut−off
r=0.7
VOCs
Fungi
Correlations
Positive Correlation
Negative Correlation
Expression
Spring
Summer
Autumn
Winter

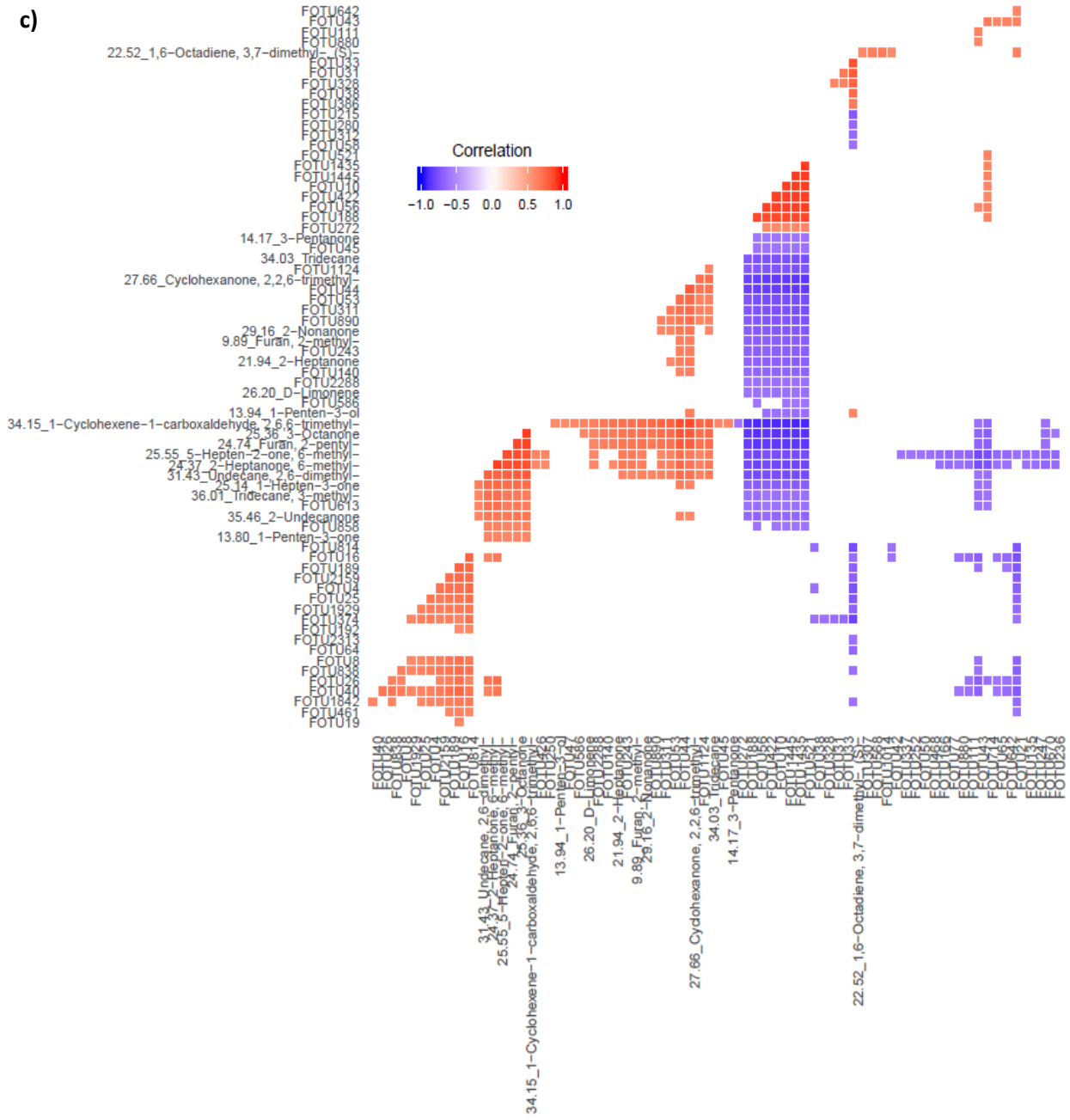


Figure 3.17 Correlation plots built from a model used to combine fungi (18s rRNA) and VOC data of faeces of horses sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). A circle plot is shown in **a** represents visualisation of positive and negative correlation between features. A circos plot to visualise specific OTU and VOC correlations generated by the circle plot is shown in **b** with a cut-off of 0.7. In **c** a heatmap showing correlations of features from the circle plot with a lower cut off (0.3 and above). Labels of OTUs and VOCs are in Appendix 3.4 and 3.2, respectively.

OTU	VOC	Corr	p-value (FDR)
OTU40	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.63	0.04
OTU10	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	-0.87	0.00
OTU56	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	-0.86	0.00
OTU1435	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	-0.88	0.00
OTU1445	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	-0.88	0.00
OTU521	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	-0.90	0.00
OTU188	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	-0.88	0.00
OTU422	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	-0.89	0.00
OTU10	1-Hepten-3-one	-0.83	0.00
OTU56	1-Hepten-3-one	-0.83	0.00
OTU1435	1-Hepten-3-one	-0.83	0.00
OTU1445	1-Hepten-3-one	-0.86	0.00
OTU521	1-Hepten-3-one	-0.83	0.00
OTU188	1-Hepten-3-one	-0.83	0.00
OTU422	1-Hepten-3-one	-0.82	0.00
OTU10	1-Propanol, 2-methyl-	0.62	0.04
OTU56	1-Propanol, 2-methyl-	0.63	0.04
OTU1435	1-Propanol, 2-methyl-	0.62	0.04
OTU1445	1-Propanol, 2-methyl-	0.64	0.03
OTU521	1-Propanol, 2-methyl-	0.62	0.04
OTU188	1-Propanol, 2-methyl-	0.63	0.04
OTU422	1-Propanol, 2-methyl-	0.62	0.04
OTU10	2-Heptanone	-0.65	0.03
OTU1435	2-Heptanone	-0.64	0.03
OTU521	2-Heptanone	-0.62	0.04
OTU422	2-Heptanone	-0.62	0.04
OTU10	2-Heptanone, 6-methyl-	-0.72	0.01
OTU56	2-Heptanone, 6-methyl-	-0.71	0.01
OTU1435	2-Heptanone, 6-methyl-	-0.73	0.01
OTU1445	2-Heptanone, 6-methyl-	-0.75	0.01
OTU521	2-Heptanone, 6-methyl-	-0.74	0.01
OTU188	2-Heptanone, 6-methyl-	-0.70	0.01
OTU422	2-Heptanone, 6-methyl-	-0.72	0.01
OTU426	2-Heptanone, 6-methyl-	0.66	0.02
OTU10	2-Nonanone	-0.74	0.01
OTU56	2-Nonanone	-0.68	0.02
OTU1435	2-Nonanone	-0.71	0.01
OTU1445	2-Nonanone	-0.65	0.03
OTU521	2-Nonanone	-0.70	0.01
OTU188	2-Nonanone	-0.70	0.01
OTU422	2-Nonanone	-0.67	0.02
OTU10	3-Octanone	-0.76	0.00
OTU56	3-Octanone	-0.74	0.01
OTU1435	3-Octanone	-0.78	0.00
OTU1445	3-Octanone	-0.79	0.00
OTU521	3-Octanone	-0.79	0.00
OTU188	3-Octanone	-0.75	0.01
OTU422	3-Octanone	-0.78	0.00
OTU1445	3-Pentanone	-0.62	0.04
OTU10	5-Hepten-2-one, 6-methyl-	-0.70	0.01
OTU56	5-Hepten-2-one, 6-methyl-	-0.69	0.01
OTU1435	5-Hepten-2-one, 6-methyl-	-0.71	0.01
OTU1445	5-Hepten-2-one, 6-methyl-	-0.71	0.01

OTU521	5-Hepten-2-one, 6-methyl-	-0.72	0.01
OTU188	5-Hepten-2-one, 6-methyl-	-0.68	0.02
OTU422	5-Hepten-2-one, 6-methyl-	-0.70	0.01
OTU40	Butanoic acid, 2-methylpropyl ester	-0.64	0.03
OTU10	Cyclohexanone, 2,2,6-trimethyl-	-0.70	0.01
OTU56	Cyclohexanone, 2,2,6-trimethyl-	-0.66	0.02
OTU1435	Cyclohexanone, 2,2,6-trimethyl-	-0.72	0.01
OTU1445	Cyclohexanone, 2,2,6-trimethyl-	-0.69	0.01
OTU521	Cyclohexanone, 2,2,6-trimethyl-	-0.72	0.01
OTU188	Cyclohexanone, 2,2,6-trimethyl-	-0.69	0.01
OTU422	Cyclohexanone, 2,2,6-trimethyl-	-0.73	0.01
OTU10	Furan, 2-methyl-	-0.71	0.01
OTU56	Furan, 2-methyl-	-0.81	0.00
OTU1435	Furan, 2-methyl-	-0.74	0.01
OTU1445	Furan, 2-methyl-	-0.75	0.01
OTU521	Furan, 2-methyl-	-0.78	0.00
OTU188	Furan, 2-methyl-	-0.77	0.00
OTU422	Furan, 2-methyl-	-0.76	0.00
OTU10	Furan, 2-pentyl-	-0.76	0.00
OTU56	Furan, 2-pentyl-	-0.75	0.01
OTU1435	Furan, 2-pentyl-	-0.78	0.00
OTU1445	Furan, 2-pentyl-	-0.83	0.00
OTU521	Furan, 2-pentyl-	-0.80	0.00
OTU188	Furan, 2-pentyl-	-0.77	0.00
OTU422	Furan, 2-pentyl-	-0.81	0.00
OTU56	Undecane, 2,6-dimethyl-	-0.62	0.04
OTU1435	Undecane, 2,6-dimethyl-	-0.62	0.04
OTU1445	Undecane, 2,6-dimethyl-	-0.63	0.04
OTU521	Undecane, 2,6-dimethyl-	-0.62	0.04
OTU422	Undecane, 2,6-dimethyl-	-0.62	0.04

Table 3.15 Correlations between anaerobic fungal OTUs (18S rRNA gene) and VOCs isolated from horse faeces. Horses were sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4).

OTU	VOC	Corr	adj p	Taxa
OTU14	1-Propanol, 2-methyl-	0.63	0.04	Pilobolus
OTU77	1-Penten-3-one	-0.66	0.02	Chytridiomycota
OTU65	1-Penten-3-one	-0.70	0.01	Blumeria
OTU858	1-Penten-3-one	0.62	0.04	Lectera
OTU140	1-Penten-3-ol	0.74	0.01	Aureobasidium
OTU890	1-Penten-3-ol	0.64	0.03	Hypocreales
OTU71	1-Penten-3-ol	0.60	0.05	Fusarium
OTU312	1-Penten-3-ol	-0.61	0.04	Sordariomycetes Incertae Sedis
OTU44	3-Pentanone	0.69	0.01	Acremonium
OTU38	1,6-Octadiene, 3,7-dimethyl-, (S)-	0.70	0.01	Tremellales
OTU4	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.76	0.00	Helotiales
OTU33	1,6-Octadiene, 3,7-dimethyl-, (S)-	0.80	0.00	Sordariales
OTU374	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.71	0.01	Colletotrichum
OTU16	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.62	0.04	Plectosphaerella
OTU1014	1,6-Octadiene, 3,7-dimethyl-, (S)-	0.62	0.04	Pleurostomophora
OTU1929	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.71	0.01	Hypsizygus
OTU189	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.64	0.03	Verticillium
OTU31	1,6-Octadiene, 3,7-dimethyl-, (S)-	0.80	0.00	Tremellales
OTU814	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.70	0.01	Colletotrichum
OTU328	1,6-Octadiene, 3,7-dimethyl-, (S)-	0.65	0.03	Tremellales
OTU25	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.75	0.01	Cyphellophora
OTU58	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.70	0.01	Sordariomycetes
OTU191	1,6-Octadiene, 3,7-dimethyl-, (S)-	0.64	0.03	Capnodiales
OTU312	1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.69	0.02	Sordariomycetes Incertae Sedis
OTU77	2-Heptanone, 6-methyl-	-0.71	0.01	Chytridiomycota
OTU44	2-Heptanone, 6-methyl-	0.71	0.01	Acremonium
OTU44	Furan, 2-pentyl-	0.68	0.02	Acremonium
OTU140	1-Hepten-3-one	0.69	0.01	Aureobasidium
OTU44	1-Hepten-3-one	0.72	0.01	Acremonium
OTU243	3-Octanone	0.64	0.03	Diaporthe
OTU44	3-Octanone	0.70	0.01	Acremonium
OTU26	5-Hepten-2-one, 6-methyl-	0.63	0.04	Sarocladium
OTU243	5-Hepten-2-one, 6-methyl-	0.61	0.04	Diaporthe
OTU77	5-Hepten-2-one, 6-methyl-	-0.68	0.02	Chytridiomycota
OTU44	5-Hepten-2-one, 6-methyl-	0.71	0.01	Acremonium
OTU33	Cyclohexanone, 2,2,6-trimethyl-	0.61	0.04	Sordariales
OTU31	Cyclohexanone, 2,2,6-trimethyl-	0.62	0.04	Tremellales
OTU44	Cyclohexanone, 2,2,6-trimethyl-	0.63	0.04	Acremonium
OTU243	Undecane, 2,6-dimethyl-	0.63	0.04	Diaporthe
OTU243	Tridecane	0.64	0.03	Diaporthe
OTU44	Tridecane	0.64	0.03	Acremonium
OTU140	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.61	0.04	Aureobasidium
OTU890	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.62	0.04	Hypocreales
OTU243	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.65	0.03	Diaporthe
OTU44	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.73	0.01	Acremonium
OTU77	Tridecane, 3-methyl-	-0.61	0.04	Chytridiomycota
OTU890	Furan, 2-methyl-	0.63	0.04	Hypocreales
OTU243	Furan, 2-methyl-	0.68	0.02	Diaporthe

OTU44	Furan, 2-methyl-	0.65	0.03	Acremonium

Table 3.16 Correlations between facultative anaerobic fungi (FF) and VOCs isolated from horse faeces. Horses were sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4).

3.3.4.2 VOCs and ITS1

The performance of the cross-validation of the VOCs and ITS1 model, displaying the overall classification error rate and BER is shown in Figure 3.18. From this the model selected the first four components (as the BER continued to decrease until this point) and the centroids distance which overall had the lowest BER was selected to reduce the chance of overfitting. Figure 3.19 shows component 1 and component 2 of a Pearson's correlation plot to show how well fungal OTUs and VOCs can discriminate between time points (Figure 3.19a and b). VOCs were superior to fungi for the separation of samples according to time point (Figure 3.19c and d). The loadings for component 1 are shown in Figure 3.20 a (fungal OTUs) and b (VOCs). From the Pearson's plot there were no significant correlations between OTUs and VOCs identified. Using the circosPlot() function positive correlations (>0.72) were observed between four anaerobic fungal OTUs (OTU21, OTU35, OTU39 and OTU45) and three esters (butanoic acid, 1-methylethyl ester; propanoic acid, butyl ester and pentanoic acid, butyl ester). The same four OTUs also negatively correlated (-0.70) with 5-hepten-2-one, 6-methyl-. OTU5 (Neocallimastigales) positively correlated with phenol, 4-ethyl- and 1H-pyrrole-2,5-dione, 3-ethyl-4-methyl- and butanoic acid, 3-methyl-, propyl ester. OTU028 (Neocallimastigales) negatively correlated with 2-heptanone, 6-methyl- (-0.76) and 5-hepten-2-one, 6-methyl- (-0.72). In terms of FF, one VOC (beta-pinene) positively correlated with OTU40 and OTU44. There were 41 VOCs that negatively correlated with FF OTUs, 19 were esters and a full list of these correlations are shown in Appendix 3.7.

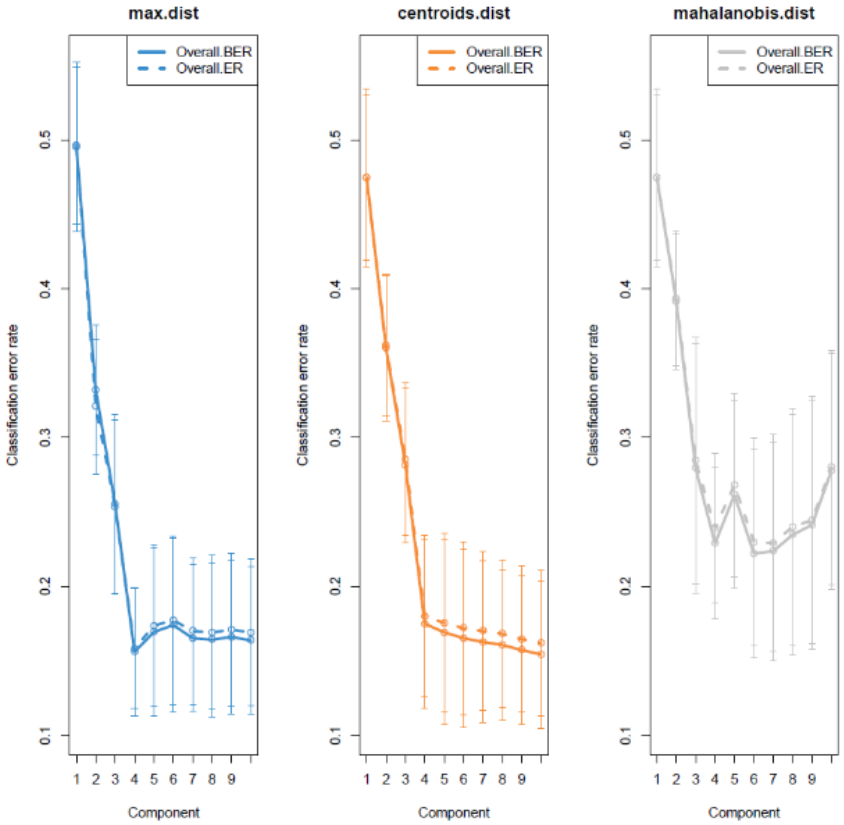


Figure 3.18 The classification error rate of a model used to combine fungi (ITS1) and VOC data of the faeces of horses sampled over four time points. Key: ER = error rate, BER = balanced error rate.

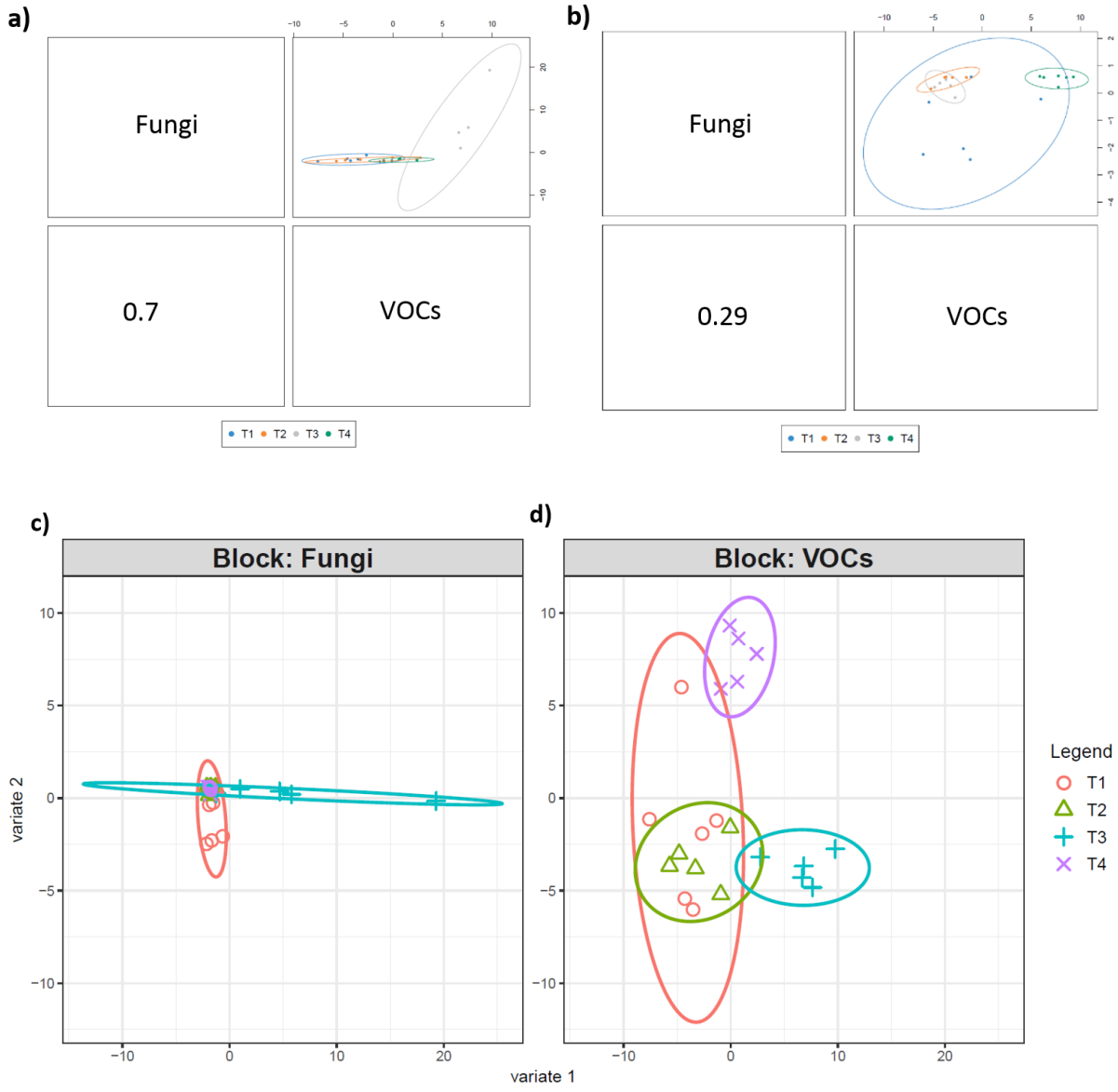


Figure 3.19 A Pearson's correlation plot of fungi (ITS1) and VOC data of the faeces of horses sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). Component 1 is shown in **a** and component 2 is shown in **b**. The ability of the model to separate time points by fungi alone is demonstrated by plot **c** and VOCs alone by plot **d**.

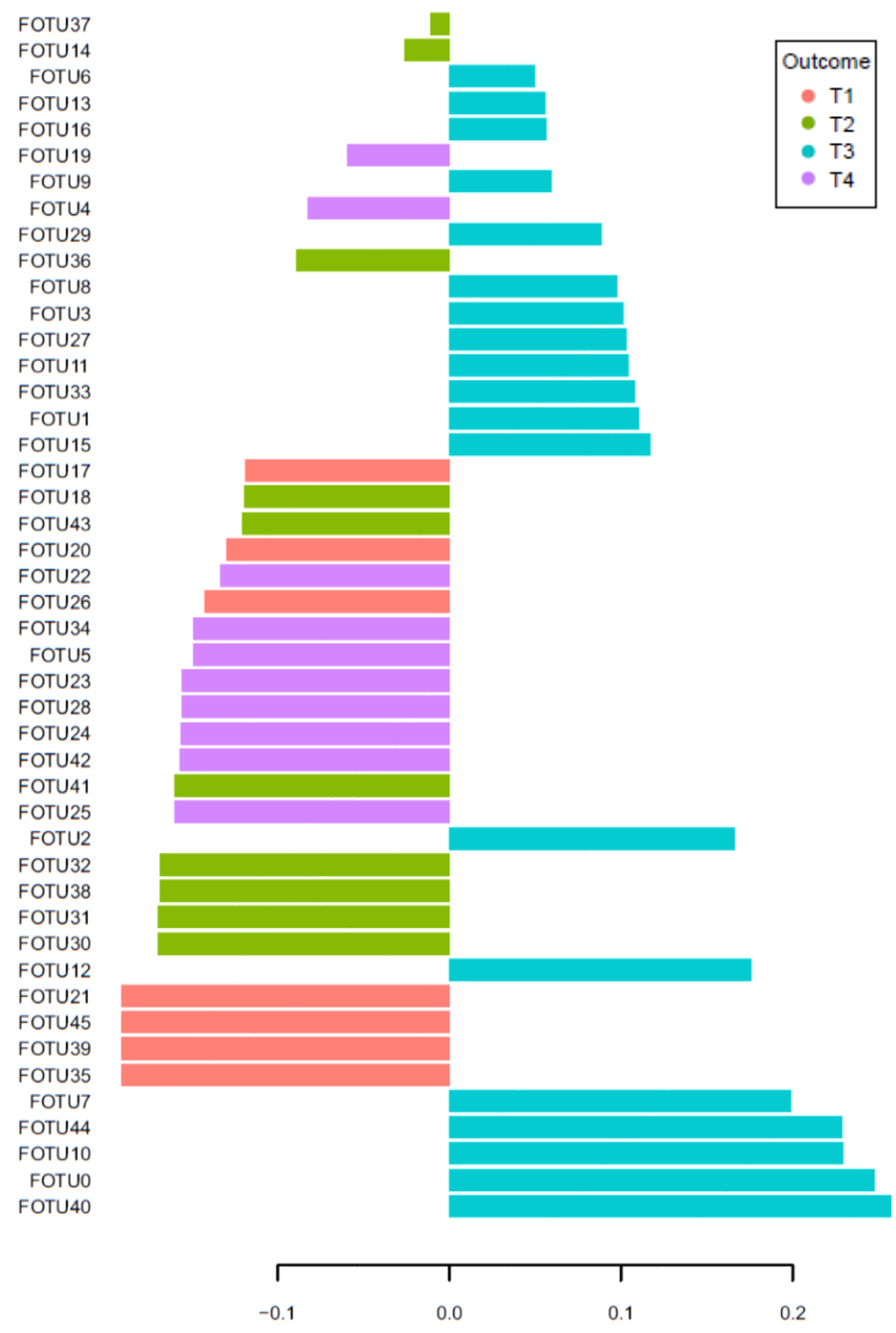


Figure 3.20a Loadings for component 1 of the Pearson's correlation plot (Figure 3.19a) of fungi of the faeces of horses sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). OTU labels can be found in Appendix 3.8.

Figure 3.20b Loadings for component 2 of the Pearson's correlation plot (Figure 3.19b) of VOCs of the faeces of horses sampled over four time points representing spring (T1), summer (T2), autumn (T3) and winter (T4). OTU labels can be found in Appendix 3.6.

3.4 Discussion

3.4.1 Reference solution

During GCMS analysis of samples, a reference solution was included weekly to check for variations in known amounts of compounds run on the GCMS. Stability of this reference solution gives greater confidence that machine bias was not introduced during analysis. The compound 2-pentanone was a major influence in separating for the batch of reference solution, which also increased once weeks 1 and 2 were removed. The variation of 2-pentanone between batches was probably because of pipetting error rather than machine bias as the other compounds were stable and 2-pentanone was stable in the reference solution in **Chapter 4**. Furthermore, the CoV of 2-pentanone remained within the threshold of 30% considered acceptable for biological studies (Drabovich *et al.*, 2013). In a PCA including all data for the faecal samples collected in this study, some clustering of samples for week of running was evident (Figure 3.1a). Post-removal of week 1 and 2 data, clustering for GCMS runtime was less apparent (Figure 3.1b). However, to adjust for any remaining runtime effect, averages of technical replicates were taken. In this thesis, checks were made for week-by-week effect of GCMS runtime and any events which may have introduced machine bias (for example re-setting of the MS) were recorded. Moreover, it is uncertain whether runtime effects may have occurred within a shorter timeframe than within-week, for example inter-day variation. In other work in this thesis (**Chapter 5**), inter-day runtime variation was investigated and was not apparent; therefore, it was not investigated here.

3.4.2 The temporal faecal VOC metabolome

In a PCA analysis of all samples, clustering was not evident for individual any horse and PERMANOVA analysis revealed that horse accounted for 10% of variation in the VOC profile. In contrast to the gut microbiota, the metabolome appears to be more conserved between individuals and it has been found that up to 33% of variation in the gut microbiota can be explained by the individual compared to 12% of variation for the metabolome (Salem *et al.*, 2019). It has been suggested by others that the equine gut microbiota may be diverse between individuals but a common functionality of these bacterial species may exist, providing a similar metabolome between individuals (Blackmore *et al.*, 2013).

The variable accounting for the most variation in the faecal VOC profile as identified by PERMANOVA analysis was time point (Table 3.11). The PCA analysis also supported this as clusters for time point were evident (Figure 3.4c). The majority of VOCs significantly

associated with change over time were esters and aldehydes. It was noteworthy that esters also increased with time as in **Chapter 2**, in faeces from one pony and esters were associated with storage at -20°C. In the current study, samples were stored at -80°C until all samples had been collected and were later transferred to a -20°C freezer prior to analysis. Because samples were transferred to a -20°C freezer at the same time, it seems unlikely that esters were a storage artefact here. Furthermore, samples collected at the April time point and were in storage for the longest and had fewer relative proportions of esters than May or July samples (Figure 3.3 Stacked plot). In PCA analysis (Figure 3.4c) the April samples also clustered away from the May-Aug samples towards Oct-Mar, most likely because of a biological impact rather than storage impact on VOC profile.

Whether the horses were consuming a diet of grass, or grass with haylage, had an important influence on the VOC profile. VOCs in the faeces may arise directly from undigested plant material consumed (Morgan & Pereira, 1962). The chemical composition of grass and conserved forage is distinct (French *et al.*, 2000). In addition, the change in VOCs associated with diet may also have been attributed to a change in substrate made available to the gut microbiota (Tang *et al.*, 2019). The three main VFAs (acetic acid, propanoic acid and butanoic acid) produced by the gut microbiota that contribute towards the energy requirements of the horse were detectable using the GCMS platform in this study. However, relative changes in these main VFAs associated with the variables recorded (e.g. feed type, temperature and season) were not observed. This is consistent with other targeted metabolome studies which have found that the main VFAs remain stable with time and across varied diets (Daly *et al.*, 2012; Blackmore *et al.*, 2013; Morrison *et al.*, 2018). A strength of this untargeted metabolome study was that a broad range of metabolites were detected that have not previously been studied over such a long period of time in horse faeces. A large number of the metabolites that altered in association with dietary change may have been emitted from bacteria (Kai *et al.*, 2009; Effmert *et al.*, 2012a). In the bacterial microbiome analysis of samples from the current work, Fibrobacteres and Spirochaetes increased in relative abundance, whilst Firmicutes decreased, after haylage was introduced at the end of December (Salem *et al.*, 2018). The shift in VOC profile appeared to mirror these changes indicating that VOCs may be used as potential markers of gut bacteria. To further strengthen the hypothesis that VOCs can act as markers for bacterial change, integration of the two omics is needed. Moreover, a noteworthy observation in this study was that a number of VOCs associated with feed type correlated to fungal OTUs in the integrated analysis as discussed further below.

A linear association between VOC profile and the average environmental temperature to which horses were exposed was a notable outcome of this study. The chemical composition of grass is known to fluctuate throughout the year in the UK (Owen *et al.*, 1978; Sarkijarvi *et al.*, 2012), which coincides with changes in the ambient environmental temperature. Chemical analyses of grass, forage and soil samples were not performed here, but it would have been useful to correlate with VOC changes with time point, diet and weather parameters. Furthermore, during the winter when temperatures are lower, grass is unable to grow so conserved forages are often fed to supplement horses' diets. In this study haylage was made available to the horses at the times of year when grazing was in short supply. Therefore, the relationship between temperature and VOC profile may be indirect, because the feed type changed with temperature. On the other hand, the effect of temperature may have had an impact on faecal VOCs once the samples had been collected. Samples were frozen within 2-6 hours of collection. In **Chapter 2**, it was shown that 98-95% of VOCs were shared with a 1 hour-frozen sample when samples were stored after 4 and 8 hours, respectively. However, the investigation in **Chapter 2** was carried out at one time point only and the effects of time from collection to freezing on VOCs at other times of the year were unknown. From the current study design, it was not possible to ascertain which factors of those recorded were responsible for changes in VOC profile. It was shown here that changes over time were evident and that feed type, temperature and season were associated with VOC profile changes. Further studies of a crossover design focusing specifically on diet or management changes are required to further understand the effect of these factors alone on faecal VOCs.

The use of faecal VOCs as biomarkers for colic has been investigated previously in a pilot study, which identified markers of potential interest (Turner *et al.*, 2013). Future studies investigating faecal VOCs as biomarkers for equine gastrointestinal disease should consider the time of year and the diets horses are consuming in the experimental design as these factors appear to influence the VOC profile of the healthy horse.

3.4.3 Relationship of FEC and VOC profile

FEC is an indicator of the presence of strongyle and ascaris parasites in the gut of live horses. Establishing relationships between parasites, the gut microbiota and the functional metabolome may help to discover new biomarkers and understand host-microbiota-parasite relationships to develop alternative drug therapies (Peachey *et al.*, 2017; Brosschot & Reynolds, 2018). A relationship between FEC and VOCs was not evident in this work. PCA and

PERMANOVA demonstrated stronger effects between time points sampled rather than for FEC category. It was not surprising that VOCs and FEC did not correlate as previous microbiome studies did not identify any major shifts in bacterial diversity with FEC; therefore it appears VOC profiling mirrors previous findings of microbiota studies (Clark *et al.*, 2018; Peachey *et al.*, 2018). FEC is not representative of real-time worm burden which may explain a lack of correlation of FEC with the microbiota and metabolome (Murphy & Love, 1997). Further studies with accurate determination of parasite burdens are required to establish whether the gut microbiome and metabolome interact with parasites in the horse. Moreover, FEC was recorded on just two occasions in this study and on the second occasion VOC analysis was performed on a sample collected 14 days later so the FEC may have altered by the time the VOC sample was collected. The application of FECs at each time point was not possible during this study because of time constraints but it would have provided a larger sample size to enable stronger conclusions to be made of the relationship between FEC and VOC profile. During the sampling period of this study horses were administered moxidectin and praziquantel 10 days prior to the samples being collected at time point May 2. It has been demonstrated that up to 14 days post-anthelmintic treatment (ivermectin specifically) a shift in microbial richness can occur (Peachey *et al.*, 2019). Therefore, it cannot be excluded that the microbiome and consequently the metabolome of the May 2 sample was influenced by a previous anthelmintic treatment.

3.4.4 The temporal faecal mycobiome and metabolome

Current knowledge about the equine hindgut mycobiome is limited. The VOC analysis in this work revealed a marked increase in potential fungal compounds in the autumn (Effmert *et al.*, 2012a). Therefore, a subset of samples was further analysed to characterise the faecal mycobiome to see if the results mirrored the VOC profile. It was hypothesised that fungi from the diet and environment (e.g. from decomposing leaves, contaminated pasture) were responsible for the spike in fungal compounds observed in the autumn. In order to detect both FF and AF, two sets of primers were selected. An 18S rRNA primer set was chosen based on the previous optimisation of this set to detect fungi (Liu *et al.*, 2012). An ITS1 primer set was chosen to target anaerobic species specifically (Tuckwell *et al.*, 2005).

The ITS1 primer set was able to detect more AF fungal reads than the 18S set. A total of 23 AF OTUs and 892,379 reads were detected by ITS1 and 10 AF OTUs and 524,358 reads were identified 18S rRNA. The ITS1 set was also able to identify AF at a higher resolution with some genera identifications made; however over 36% of AF sequences could not be identified at

genus level. A large number of unclassified sequences in ruminant ITS1 based studies has been commonly reported and may be because ITS1 databases are not as developed as bacterial databases (Kumar *et al.*, 2015). Several other clades including *AL1, AL7, DT1, KF1* from the *Neocallimastix* genus and other genera including *Orpinomyces* and *Anaeromyces* have been previously identified in the faeces of a horse but were not detected in this work (Mura *et al.*, 2019). The ITS1 region targeted by the primer set here may not have been conserved between organisms (Edwards *et al.*, 2008). Therefore, a primer set applied by the recent study performed by Mura *et al.* (2019) using a forward universal ITS1 primer and a reverse 5.8S rRNA gene specific to Neocallimastigomycetes may have been more appropriate. The ITS1 primer set selected here was based on ruminant studies as published molecular-based studies characterising anaerobic fungi of horse faeces were very limited when this experiment was planned and carried out (Liggenstoffer *et al.*, 2010; Tapio *et al.*, 2017).

As predicted from the VOC results, the mycobiome in the autumn showed that the relative abundance (and OTU counts) of AF decreased and abundances of FF increased. The diversity in the mycobiome characterised by both primer sets also increased at the autumn time point. Many of the FF (members of order Hypocreales) which increased during the autumn were plant pathogens or species known for growing on plants, in soil or in decomposing leaf litter (Effmert *et al.*, 2012a; De Melo *et al.*, 2018). Fungal plant pathogens tend to favour wet and mild conditions, which is characteristic of autumn in the UK (West *et al.*, 2012). It is unclear why AF and FF had an inverse relationship here. Studies characterising AF and FF in the horse are limited and, to date, there is no published study reporting an interaction between the two (Doxey *et al.*, 1990). Statistical modelling in the current work demonstrated, for the first time, positive and negative correlations between AF and FF OTUs in horse faeces (Figure 3.18c). Further investigation of the relationship between AF and FF in the faeces of the horse is warranted because this event coincides with an increased incidence of colic in the autumn (Hillyer *et al.*, 2001; Archer *et al.*, 2006). It could be suggested that primer amplification bias (including mismatches, annealing temperature) towards a certain species may have had a role in the differences observed (Sipos *et al.*, 2007). Preferential selection of smaller sized ITS1 regions has recently been reported for ITS1 primers targeting mock communities of anaerobic fungi (Edwards *et al.*, 2019). Therefore, there are limitations in the use of the ITS1 region as a marker for anaerobic fungi. The mycobiome study performed in this thesis was a preliminary investigation and further work obtain better markers for anaerobic fungi is needed to confirm the results obtained here.

The ITS1 primer demonstrated that the increase in FF and decrease in AF in the autumn was observed in just two animals which were of Welsh pony type. Welsh ponies are a native breed and are adapted to survive in harsh mountain conditions. Native ponies have been reported to alter their browsing habits according to seasonal availability of food (Putman *et al.*, 1987). It may be possible that the Welsh ponies in this study consumed a wider range of plant material during the autumn as an adaptation of their breed to seasonal changes. This can only be speculated as horse behaviour and intake were not recorded. Moreover, the unique fungal profile in the autumn observed in the Welsh ponies may have been an adaptation of the resident microbiota to the invading organisms. As discussed in **Chapter 1**, there is some evidence that horse genetics and breed can impact the gut bacteria, although the impact of genetics on resident gut fungi is yet to be investigated. The bacterial microbiome work of these samples did not observe any obvious differences between the breeds of horses, but it did demonstrate strong individual trends (Salem *et al.*, 2018).

This is the first work to characterise the equine faecal mycobiome longitudinally using a next-generation sequencing method. The sequencing methods used here allow identification of organisms and relative abundances to be compared but it was not quantifiable. Longitudinally the concentration of AF has previously been investigated in horse faeces, with a general decrease (statistics were not performed) in concentration when horses switched to a pasture from a hay-based diet (Clark *et al.*, 2018). This pattern agrees with the current work that AF OTUs were higher drivers of separation at T1 when the horses had been consuming a conserved forage diet over the previous winter (Figures 3.17a and 3.21a). Building on the current work, specific taxa could be selected and quantified using qPCR to support the inverse relationships shown here.

In common with studies of human gut fungi, it is difficult to determine here whether FF in the faeces were simply inactive contaminants from the diet or whether they can colonise and have an influence on the rest of the gut ecosystem (Rizzetto *et al.*, 2014). A proportion of VOCs detected in this study may be produced by fungi (listed in Tables 3.15 and 3.16) and include 3-octanone; 5-hepten-2-one, 6-methyl-; 2-heptanone, 6-methyl-; 2-heptanone and 1-penten-3-ol (Combet *et al.*, 2006; Effmert *et al.*, 2012a). A number of these VOCs positively correlated with FF so this may suggest they are capable of being active in the gut and therefore have colonised it. The consequences of metabolites produced by FF on the rest of the gut ecosystem and the host remains unknown. In humans, FF (including *Candida albicans*) have been linked with the modulation of host immune responses and intestinal inflammation (Jawhara *et al.*, 2008; Zhang *et al.*, 2017). In horses the presence of FF in the hindgut and the

role of FF in health and disease has received very little attention (Doxey *et al.*, 1990). In cattle one study reported a lower fungal diversity in faecal samples of those diagnosed with clinical mastitis compared to healthy controls (Wen *et al.*, 2018). From these studies and the current work, the hypothesis that FF may play a role in of the development of intestinal inflammation in the horse deserves further investigation.

Numerous studies have identified faeces as a suitable proxy for studying the bacteria of the latter regions of the horse hindgut (Dougal *et al.*, 2012; Costa *et al.*, 2015a; Fliegerova *et al.*, 2016). In terms of AF, a study in one horse revealed all species identified in the latter hindgut regions were also present in the faeces, but differed considerably in terms of relative abundance and concentration along the tract. Further work in larger sample sizes and multiple sampling time points is necessary to determine whether faeces are a suitable proxy for studying hindgut AF in the horse.

From this work, it appears that VOCs have the potential to be markers of fungal populations in horse faeces. Notably, in both primer sets 5-hepten-2-one, 6-methyl- and 2-heptanone, 6-methyl-correlated negatively with anaerobic fungal OTUs and may be produced by a fungal plant pathogen (Demyttenaere & De Kimpe, 2001). It may be reasonable to speculate that some of the compounds produced by FF or metabolites arising directly from the plants themselves consumed during the autumn may have deleterious effects on anaerobic fungi. When studied *in vitro,* low concentrations of secondary plant metabolites (p-coumaric, ferulic and sinapic acids) decreased AF populations (Akin & Rigsby, 1985). Despite the significant fluctuations in diversity and abundance of fungal species observed here, horses remained healthy throughout the sampling period, so it can be presumed patterns were within the range of healthy horses. Furthermore, metabolites previously identified as being produced by AF (including ethanol and acetic acid) did not correlate with AF changes, indicating that the overall function of the AF population remained stable (Cheng *et al.*, 2013).

3.4.4 The core faecal metabolome and mycobiome

Over the 12-month study, 63% of compounds were observed at all time points in at least one horse. This further supports the hypothesis of a more conserved faecal metabolome in the horses compared to the microbiome. Many of the core compounds belonged to classes of acids, esters, aldehydes and alcohols and have previously been associated with microbial fermentation in ruminants (Cai *et al.*, 2006). In work by Dougal *et al.*, (2017), the core faecal bacteria in horses maintained on uniform diets for 6 weeks were reported to represent 21-28% of total sequences. Here, core fungal OTUs represented 63% and 82% for 18S rRNA and

ITS1 primer sets, respectively. This could be interpreted that the fungal microbiome of the horse may be more conserved than the bacterial microbiome. However, the primer sets used here to study the mycobiome have limited or no application to studying both FF and AF in horse faeces; therefore, it is highly likely more species than were detected here are actually present. Further studies to optimise extraction and primer choice to study the fungal mycobiome of the horse are required before accurate estimates of the core organisms can be made.

3.4.5 Overall discussion and conclusions

Over a 12-month sampling period the faecal VOC profile of grazing horses altered. As identified in statistical modelling, the factors average temperature, diet and time point had an effect on the greatest number of VOCs. A second year of sampling of the same group of horses would strengthen the results and conclusions made here. As mentioned, a crossover designed study focusing on a single factor of interest e.g. diet or temperature would better explain the impact of each factor. The feasibility of a controlled trial for an early study in equine faecal VOCs was not possible because of the costs and labour involved. The results here were in agreement with a two-year study of the microbiota of wild mice, that seasonal changes were also likely because of changes in diet associated with season (Maurice *et al.*, 2015). Further supporting this, a primate study also found that rainfall and diet were most associated with seasonal changes (Orkin *et al.*, 2018). For further confirmation that some of the VOC changes observed here were microbiota related, integration with previously produced 16S rRNA data is needed as a next step, which was unfortunately beyond the timescale of this thesis. Nevertheless, for the first time VOCs have been shown to be potential markers of fungi in the faeces of grazing horses. This is a first step towards the development of a simple, cost-effective and rapid tool for studying the stability of equine hindgut microorganisms. Further work in health and disease should build on this and carefully consider changes in season and diet when designing experiments.

Chapter 4 Temporal variation in the faecal VOCs of the periparturient mare

4.1 Introduction

The broodmare is at an increased risk of colic and is 13-times more likely to suffer from large colon volvulus than the stallion or gelding (Suthers *et al.*, 2013). The most common types of colic that occur in broodmares are large colon displacement, large colon torsion, epiploic foramen entrapment, other small intestinal strangulating lesions and ileal impaction (Suthers *et al.*, 2014). The precise mechanism underlying an increased risk of colic in the broodmare is unknown. Physical damage caused by the presence of a foetus or the process of parturition may be responsible for certain types of colic including caecal ruptures or mesenteric tears. It has also been speculated that a change in management factors around time of parturition or a sudden increase in space in the abdomen due to the reduced volume of the uterus and its contents may result in altered gastrointestinal motility and positioning. There are great hormonal and immunological changes that occur during pregnancy. In the human the elevated level of progesterone acting on the gut is believed to be a common cause of constipation during pregnancy (Marshall *et al.*, 1998). However, the impact of hormones on the equine gut during pregnancy is unknown.

In particular, the periparturient mare is at high risk of colic during the first 90 days post-foaling (Suthers et al. 2013). Physiological changes (increased non-esterified fatty acids) in the pre and post-partum mare may interact with gut microbiota and induce changes that may directly or indirectly cause colic (Koren *et al.*, 2012; Holcombe *et al.*, 2014). Large colon displacement and large colon torsion have both been linked to changes in the microbiota (Daly *et al.*, 2012). Previous work has shown that the microbiota of the healthy mare does not change in response to parturition (Weese *et al.*, 2014). However, in the same study, mares which developed colic during the periparturient period had a lower abundance of Firmicutes and a greater abundance of Proteobacteria preceding the episode, when compared to healthy mares.

To date, there has been no work to characterise the faecal VOC metabolome around the time of parturition in the mare. The metabolic changes that occur during pregnancy may cause the normal fluctuations of the faecal metabolome of the broodmare to differ from stallions, geldings and barren mares; therefore, it is important to study this category of horse individually. This information would provide a benchmark for the fluctuations that occur

normally at this time and deviations from this may represent a disease state. In this chapter the change in the faecal VOC metabolome of the broodmare was studied 5 weeks before foaling, during second stage labour and 12 weeks post-foaling.

4.2 Methodology

4.2.1 Sample collection

Faecal samples were collected from seven mares at a stud farm in the North West of England. Sampling took place five weeks prior to the mares foaling, during second stage labour and 12 weeks post-foaling (from mid-April to the end of July 2014). Timings of sample collection are summarised in Table 4.1. Figure 4.1 shows the management regimens of the mares over the course of the sampling period. When foals were sick, mares were stabled for 24 hours a day and this was recorded as regimen 5 (R5). The times when mares were stabled is shown in Table 4.1.

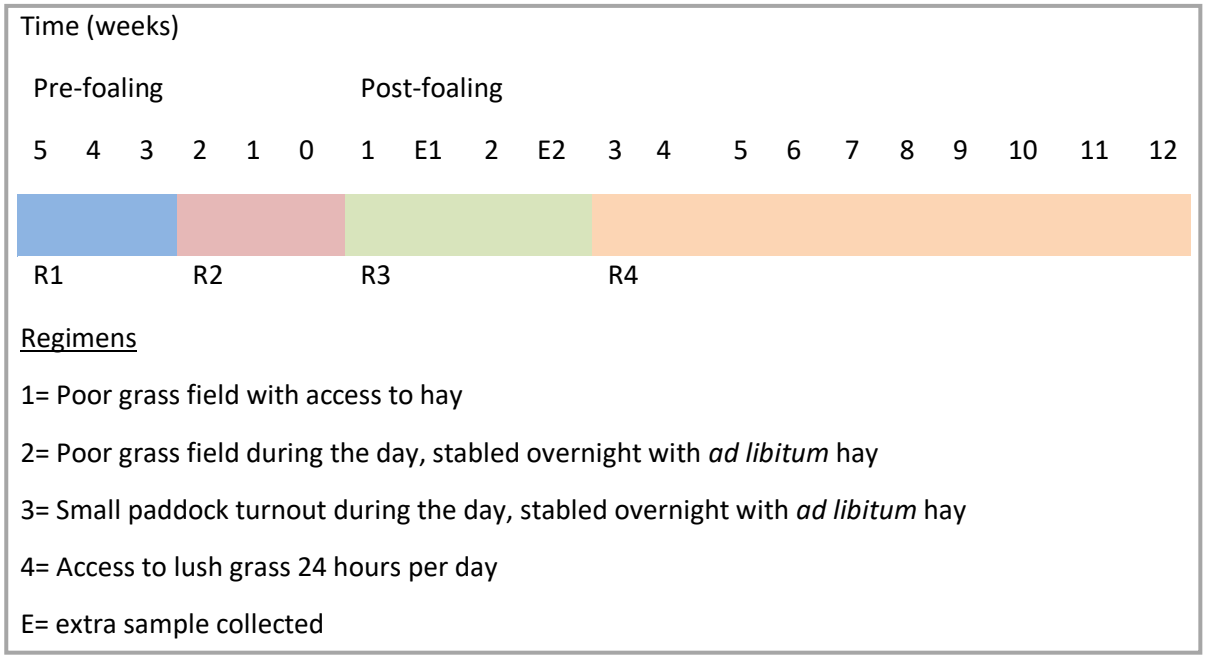


Figure 4.1 Management of mares during study period. There were some exceptions to the regimen (Table 4.1). Some horses were stabled when foals were sick (regimen 5) – refer to Table 4.1.

Time, weeks	Number of days sample was collected to foaling	Regimen	Notes
Pre-foaling			
T -5	29-33	1	NA
T -4	22-26	1	NA
T -3	14-19	1	NA
T -2	7-12	2	Horse S6 was on regimen 1
T -1	18 hours to 5 days	2	Horse S6 was on regimen 1
T0	During second stage labour	2	Horse S6 was on regimen 1
Post-foaling			NA
T1	3-7	3	Horse S4 was on regimen 4
Extra sample 1	6-6	3	NA
T2	9-15	3	Horse S4 and horse 6 were on regimen 4
Extra sample 2	14	3	NA
T3	17-21	4	Horse S5 was on regimen 3. Horse S2 was stabled 24 hours a day (regimen 5)
T4	26-28	4	NA
T5	31-37	4	NA
T6	38-42	4	Horse S1 and horse S4 were stabled 24 hours a day (regimen 5)
T7	45-49	4	Horse S4 was stabled 24 hours a day (regimen 5)
T8	52-56	4	Horse S4 was on regimen 3
T9	59-63	4	Horse S5 was stabled 24 hours a day (regimen 5)
T10	66-70	4	NA
T11	73-77	4	Horse S4 was stabled 24 hours a day (regimen 5)
T12	80-84	4	NA

Table 4.1 Exceptions to the management regimen. Regimens were as follows: 1 = poor grass field with access to hay. 2= Poor grass field during the day, stabled overnight with *ad libitum* hay**.** 3= Small paddock turnout during the day, stabled overnight with *ad libitum* hay**.** 4= Access to lush grass 24 hours per day. 5= Mares were stabled prior to sampling (foals were sick).

4.2.2 Broodmares

The mares were of similar height and breed (sportshorse type). Their ages ranged from 6 to 15 years (Table 4.2). All mares received anthelmintic administration of moxidectin 2-3 months before initial sampling began. Mares that remained in the study received a second administration of moxidectin post-foaling (the exact timing of this for each mare can be seen in Table 4.3)

Three mares were excluded from the study: horse S3 was sold two weeks after foaling, horse S6 was excluded two weeks after foaling as the foal died. Horse S7 was sold seven weeks

after foaling, but the data collected before that time were included in the final analysis. Table 4.4 describes the final sample set. Two mares received hormone treatments during the post-foaling sampling period. At 5 weeks post-foaling horse S1 was administered Ovuplant™ (deslorelin), covered by a stallion and had treatment with oxytocin for 3 days. Horse S4 was given prostaglandin and Chorulon™ (human Chorionic Gonadotropin) and was covered 6 weeks after foaling.

Mare	Breed	Age	Height (hands)	Foaling due date	Actual foaling date
S1	TB	13	16	22/05/2014	25/05/2014
S2	TBX	13	16.2	11/06/2014	15/06/2014
S3	TBX	15	16.2	20/06/2014	16/06/2014
S4	TBX	10	16.1	20/06/2014	04/06/2014
S5	WB	7	16.1	22/06/2014	20/06/2014
S6	TBX	6	16.2	28/06/2014	07/06/2014
S7	TBX	12	16.1	20/07/2014	26/06/2014

Table 4.2 Details of mares included in study Key: TB = Thoroughbred, TBX = Thoroughbred cross, WB = Warmblood.

Mare	Worm faecal egg count at start of sampling period (e.p.g)	Moxidectin administration
S1	0	2 days prior to collection of 6 week faecal sample
S2	0	2 days prior to collection of 3 week faecal sample
S3	0	N/A
S4	850	5 days prior to collection of 3 week faecal sample
S5	0	2 days prior to collection of 2 week faecal sample
S6	125	N/A
S7	0	2 days prior to collection of 1 week faecal sample

Table 4.3 Details of faecal worm egg counts and anthelmintic protocol of mares pre- and post-foaling.

Mare	-5	-4	-3	-2	-1	0	1	E1	2	E2	3	4	5	6	7	8	9	10	11	12
S1			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S2	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S3		✓	✓	✓	✓	✓	✓	✓	✓											
S4			✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
S5	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
S6				✓	✓		✓		✓											
S7	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓					

Table 4.4 Final sample set. A faecal sample was collected for each horse and is marked with a (✓). Samples highlighted in grey were taken forward for HS-SPME-GCMS. Orange (samples were stored at 4°C for up to one week) and green denotes samples following hormone or anthelmintic treatment. Orange and green samples also underwent HS-SPME-GCMS but were temporarily excluded from the analysis downstream to check for any impact they may have had on the results. Key: E = extra sample.

4.2.3 Sample preparation and GCMS analysis

Samples were transferred to the laboratory on ice and were stored at -80°C until transferred to -20°C (suitable temperature of storage for VOC work) on 25th May 2015. Samples collected during second stage labour were stored in the fridge for up to one week before being transported to the laboratory and were stored in the same way.

Samples were prepared for VOC analysis using the optimal method for extracting VOCs from horse faeces, as detailed in **Chapter 2**. Prepared samples then underwent HS-SPME-GCMS analysis (refer to **Chapter 2**). Three technical replicates were analysed for each faecal sample. The running order of samples was computer generated at random to prevent any run-time bias. A reference solution was made-up as detailed in **Chapter 2**. This was then stored in the fridge, with a shelf-life of three weeks. The same reference solution was used for week 1, 2 and 3. A new solution was prepared for weeks 4, 5 and 6 and a final solution prepared for week 7. Aliquots of 100 µl were added to 10 ml vials and were analysed using HS-SPME-GCMS and run in triplicate.

4.2.5 Identification of compounds and data formatting

Data processing and the method for averaging of technical replicates was performed as outlined in **Chapter 2**. Horse S3 and horse S6 were not included in the data analysis because they were excluded 10 weeks before the collection period ended. Time points Ex 1 and Ex 2 were also not included in the data analysis as samples were only provided by horse S1 and horse S2. GCMS data generated during the second week of the GCMS run were discarded, as

a leak was detected in the GCMS system, and these samples were outliers from the rest of the sample set. Figure 4.2a shows a PCA before removal and Figure 4.2b a PCA after removal of week 2 samples. After the removal of week 2 data, any samples with less than three technical replicates were dealt with by taking an average of the remaining two replicates, which was taken forward to the final analysis.

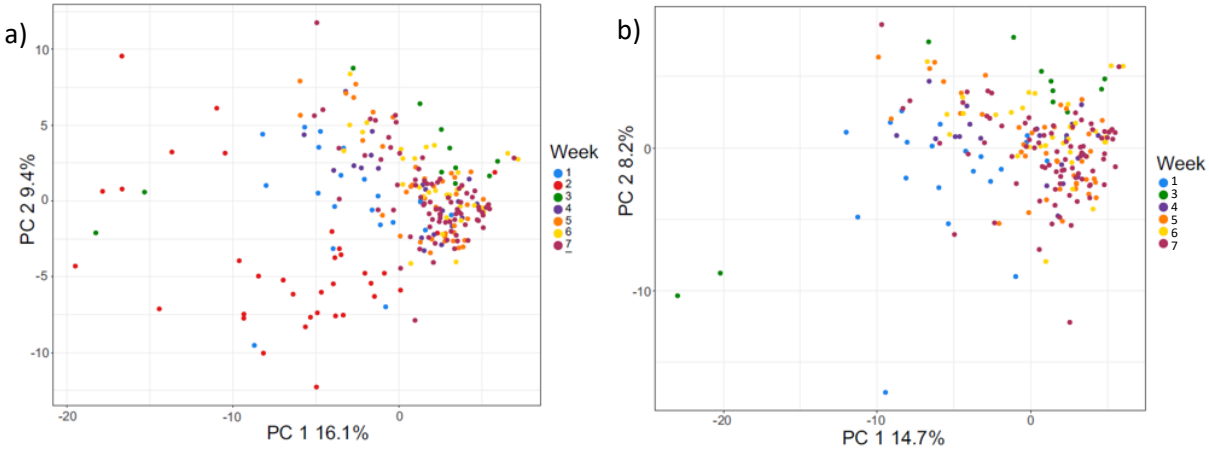


Figure 4.2 Samples grouped for week of running on the GCMS. All samples run are shown in **a).** In **b)** the data collected during week 2 was removed.

4.2.7 Statistical analysis

VOC data were processed as described in **Chapter 2**. PCAs and PERMANOVA analyses were performed focusing on the factors of time, regimen and individual horse. Linear mixed effects (LME) modelling was performed on each VOC using the R package vegan. In the LME model, time and regimen were set as fixed effects and mare was set as the random variable. P-values were FDR corrected for multiple comparisons. Statistical analysis was repeated on the data after removal of samples that had been stored in the fridge at T0 and samples that were collected a week post-anthelmintic and hormone treatments (referred to as AH samples from here on). This was to check if these factors had influenced the results.

4.3 Results

4.3.1 Reference solution

PCAs were constructed on the abundance of compounds in the reference solution samples (Figure 4.3). Indole was missing from the reference solution run in week 6 and from two out of three replicates run in week 5. For the benefit of the PCA, these missing values were replaced with half minimum value.

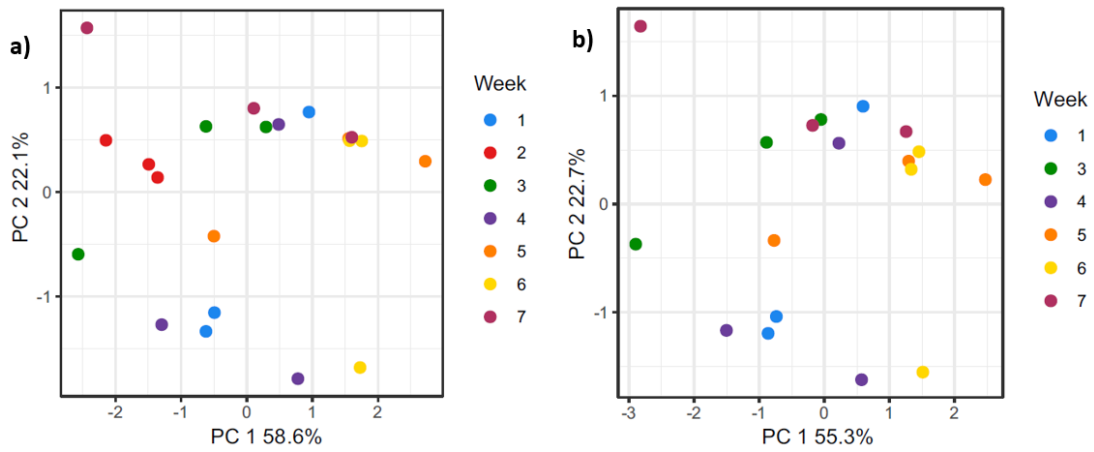


Figure 4.3 a) Abundance of compounds in reference solution samples run on weeks 1-7 and b) abundance of compounds in reference solution but without the samples run during week 2.

4.3.2 Number of compounds

A total of 105 compounds were detected in the faeces of the mares. A complete list of all compounds found is shown in Appendix 4.1. A Shapiro-Wilk test revealed that the number of VOCs detected at each week was not normally distributed, specifically at week T-4 (p=0.02). A Kruskal-Wallis test then revealed there was no significant difference in VOC numbers between weeks (p=0.56). The mean numbers of VOCs found in faeces at each week are shown in Figure 4.4.

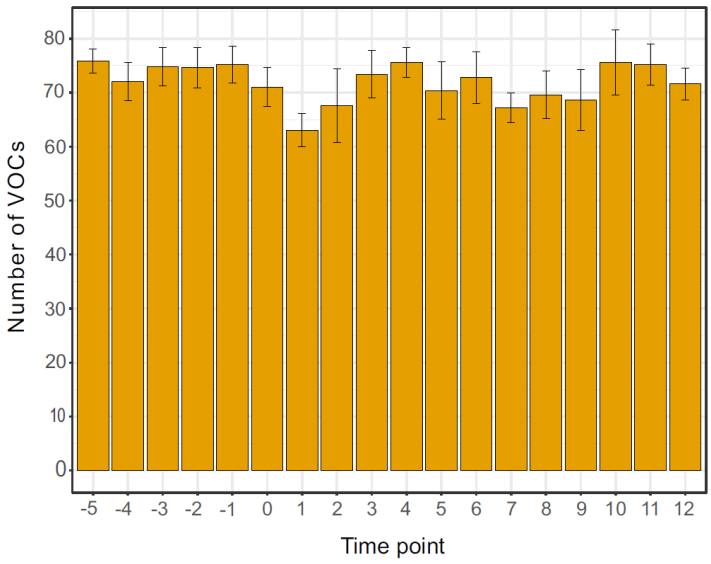


Figure 4.4 The mean numbers of VOCs. The graph illustrates VOCs emitted from the headspace of faeces as detected by HS-SPME-GCMS. Samples were collected from mares 5 weeks before foaling, during second stage labour (time point 0) and 12 weeks post-foaling. Bars represent the standard error of the mean.

4.3.3 Principal component analysis

Principal component analysis plots were constructed based on time (Figure 4.5), regimen (Figure 4.6) and the individual mare (Figure 4.7). Clustering of samples was not evident for time or mare, but some clustering was observed for regimen. The top 10 PC scores along the positive and negative axis of PC1 are shown in Table 4.5 and the complete list is in Appendix 4.1.

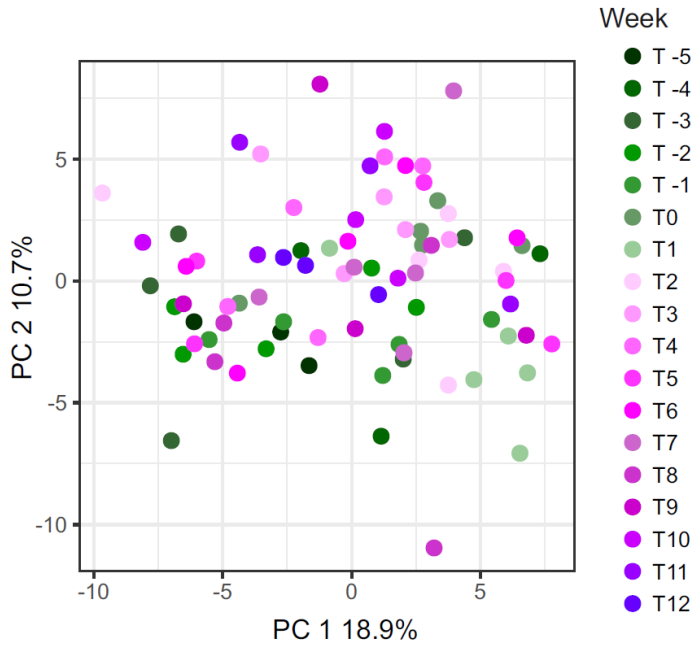


Figure 4.5 A PCA of the VOC profile of mares 5 weeks before, during second stage labour and 12 weeks post-foaling. Points are colour coded for the factor time.

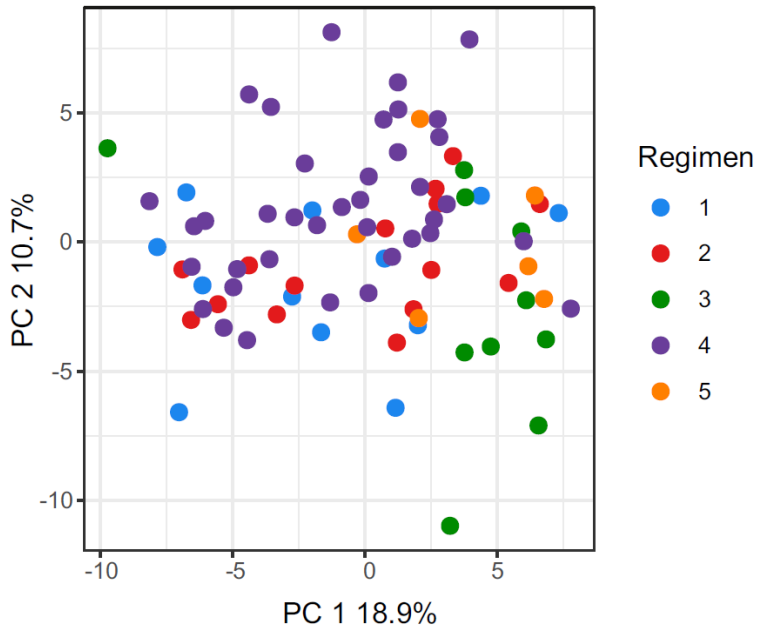


Figure 4.6 A PCA of the VOC profile of mares 5 weeks before, during second stage labour and 12 weeks post-foaling. Points are colour coded for the diet regimen. Regimens were as follows: 1 = poor grass field with access to hay. 2= Poor grass field during the day, stabled overnight with *ad lib* hay**.** 3= Small paddock turnout during the day, stabled overnight with *ad libitum* hay**.** 4= Access to lush grass 24 hours per day. 5= mares were stabled prior to sampling (foals were sick).

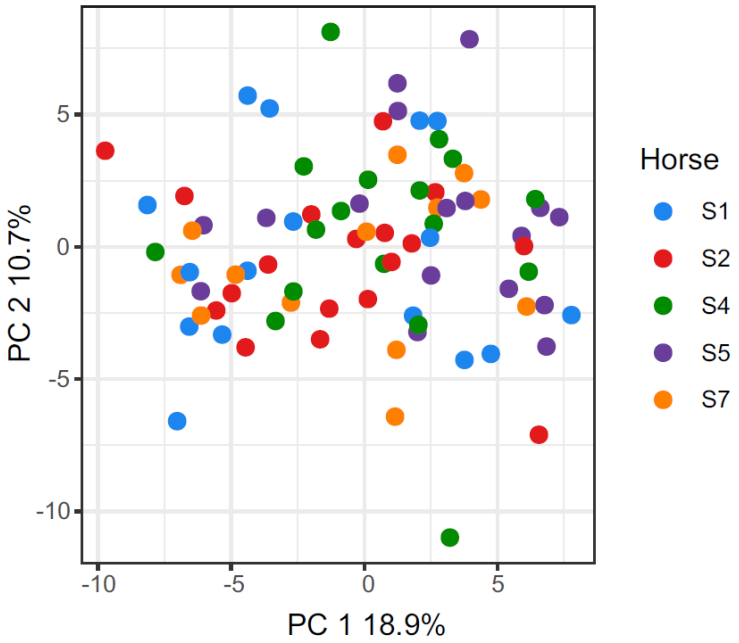


Figure 4.7 A PCA of the VOC profile of mares 5 weeks before, during second stage labour and 12 weeks post-foaling. Points are colour coded for each horse.

Negative direction		Positive direction	
VOC	**Score**	**VOC**	**Score**
Butanoic acid, 2-methylpropyl ester	-0.20	2-Butanone	0.09
Pentanoic acid, propyl ester	-0.19	Hexanal	0.08
Butanoic acid, 2-methylbutyl ester	-0.19	Benzene, 1,2,4-trimethyl-	0.05
Propanoic acid, butyl ester	-0.19	Heptanal	0.05
1-Butanol, 3-methyl-, propanoate	-0.19	3-Octanone	0.05
Butanoic acid, 2-methyl-	-0.18	Pentanal	0.05
Propanoic acid, propyl ester	-0.18	Octane	0.05
Butanoic acid, 3-methyl-	-0.18	1-Hepten-3-one	0.05
Butanoic acid, ethyl ester	-0.18	Tridecane	0.04
Propanoic acid, 2-methylpropyl ester	-0.17	2-Heptanone, 5-methyl-	0.04

Table 4.5 The top 10 highest PC scores for PC1 in both the positive and negative directions of the axis for PCAs in Figures 4.4 -6.

4.3.4 Linear mixed effects modelling

The LME model identified one compound (dimethyl sulfone) that was significantly associated with time ($p<0.001$). Line graphs of VOC concentration of dimethyl sulfone and two VOCs (nonane and 2-octene, 2,6-dimethyl-) which showed a trend towards being significantly associated with time ($p=0.07$) were plotted in Figure 4.8. Dimethyl sulfone was significantly associated with regimen. Compounds which showed a trend (but were not significant after

FDR correction) towards being associated with the factors time and regimen are listed in Table 4.6.

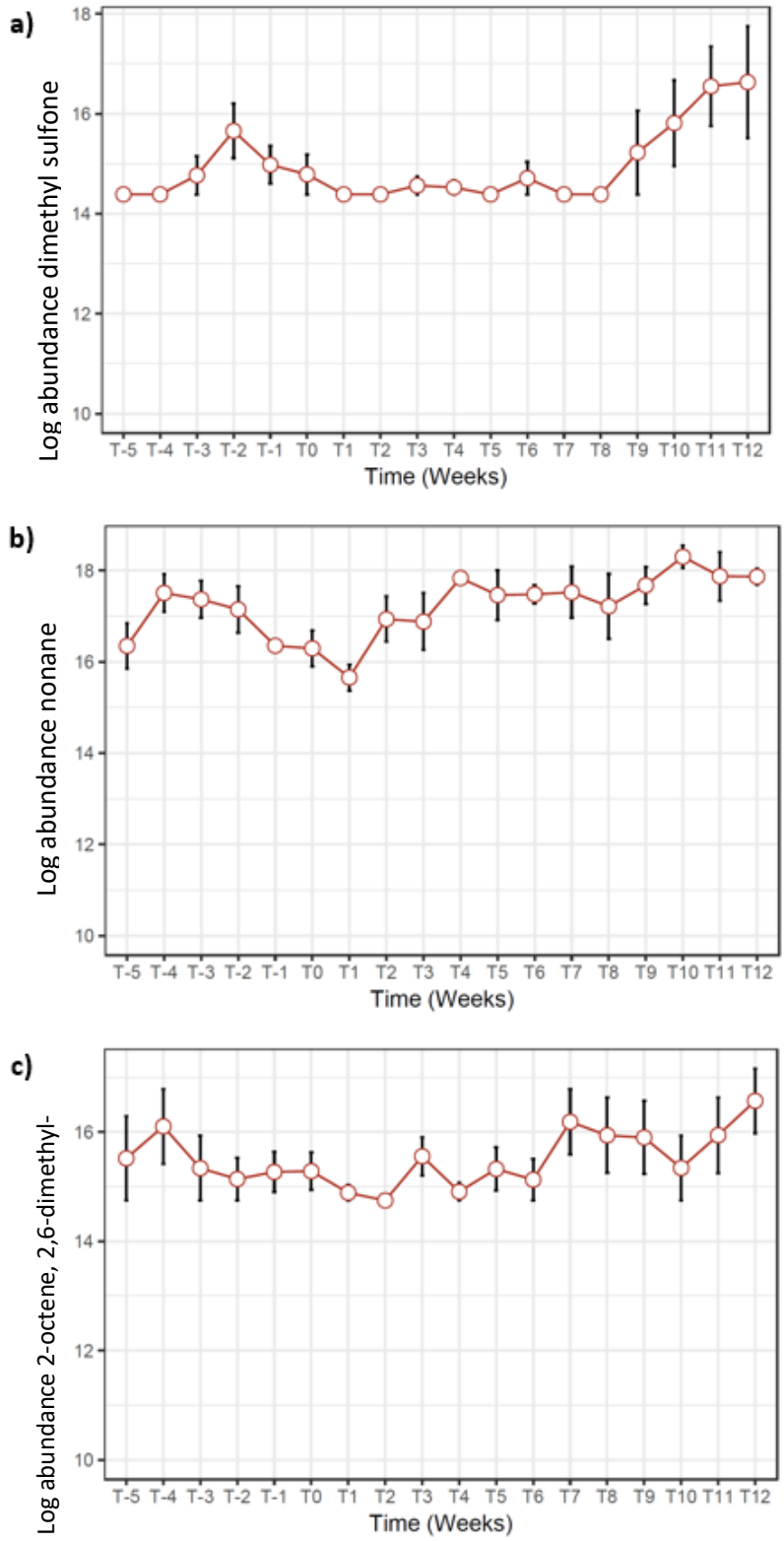


Figure 4.8 The mean log abundance of VOCs which altered significantly with time a) Dimethyl sulfone, b) nonane and c) 2-octene, 2,6-dimethyl- ($p<0.1$ LME, FDR corrected) from faecal samples collected from mares 5 weeks before foaling, during second stage labour (T0) and 12 weeks post-foaling.

VOC	*p*-value	Adjusted *p*-value
Time		
Dimethyl sulfone	<0.001	<0.001
Nonane	0.002	0.070
2-Octene, 2,6-dimethyl-	0.002	0.070
Butanoic acid, methyl ester	0.009	0.234
Pentadecane	0.013	0.273
cis-2,6-Dimethyl-2,6-octadiene	0.021	0.368
2,6-Dimethyl-2-trans-6-octadiene	0.026	0.381
D-Limonene	0.029	0.381
2-Undecanone	0.037	0.384
Butanoic acid, ethyl ester	0.038	0.384
Tridecane, 2-methyl-	0.040	0.384
Regimen		
Dimethyl sulfone	<0.001	0.042
1-Butanol, 3-methyl-, acetate	0.006	0.182
Propanoic acid, propyl ester	0.009	0.182
Butanoic acid, 2-methylpropyl ester	0.009	0.182
1-Butanol, 2-methyl-, acetate	0.010	0.182
Hexanal	0.010	0.182
Butanoic acid, ethyl ester	0.013	0.186
Propanoic acid, 2-methylpropyl ester	0.014	0.186
Butanoic acid, 2-methylbutyl ester	0.016	0.190
Dodecane, 2,6,10-trimethyl-	0.029	0.290
2-Octene, 2,6-dimethyl-	0.031	0.290
1-Butanol, 3-methyl-, propanoate	0.035	0.290
2-Undecenal	0.036	0.290

Table 4.6 Compounds that were associated with the factors time and regimen. P-values are shown for significant compounds before and after FDR correction for multiple comparisons.

4.3.5 PERMANOVA

Results obtained from a PERMANOVA analysis found that the variables time (R^2 = 0.03), regimen (R^2 = 0.03) and mare (R^2 = 0.09) were able to explain some variability in the data with *p* values of 0.006 and 0.032 and 0.003, respectively.

4.3.6 Exclusion of samples

Statistical analysis was repeated on the dataset after removal of samples that were stored at 4°C for up to a week (those collected at T0) and samples that were collected following AH treatments (Table 4.4). PERMANOVA analysis revealed that the variables time (R^2 = 0.04) and mare (R^2 = 0.10) were able to explain variation in the data with *p* values of 0.014 and 0.008, respectively. The variable regimen (R^2 = 0.03) could not explain any variation in the data (*p* = 0.06). LME modelling results and PCAs are shown in Appendix 4.2 and 4.3, respectively. Overall the exclusion of T0 and AH samples appeared to have very little impact on the results.

4.4 Discussion

4.4.1 Reference solution

Reference solution samples run during week 2 on the GCMS are clustered together in a PCA plot (Figure 4.3a). Removal of week 2 samples resulted in no distinct clustering of all three technical replicates for week of running (Figure 4.3b). This indicates that an issue occurred at week 2 (faecal samples run on the GCMS during week 2 clustered away from the rest in a PCA shown in Figure 4.2a). A leak was detected and rectified at the end of week 2 which may explain why the samples immediately before the leak detection were outliers. According to the manufacturer's handbook (PerkinElmer, 2007) air leaks in GCMS systems are a common occurrence which can interfere with optimal performance and cause poor reproducibility of results. As discussed in section 4.2.5 all faecal samples that were run during week 2 were excluded from the analysis.

Indole was not detected in 2 out of 3 samples of the reference solution at week 5 and 0 out of 3 samples at week 6 of running. Week 5 and week 6 reference solution samples were from the same batch; therefore, it is possible that pipetting error may have occurred. An alternative explanation is that the variability seen in the detection of indole when a known amount is analysed may be a limitation of the SPME fibre coating used. There is no single fibre suited to all compounds; therefore, as many fibre combinations as possible are recommended to detect all analytes. It is possible that the DVB-CAR-PDMS fibre used in this work was not the optimal for detecting indole. Previous work has found that PA and PEG are the optimal SPME fibres for detecting indole from human faecal samples using GCMS (Dixon *et al.*, 2011). It has been suggested that the late retention time (38.57 minutes) of indole in the chromatogram coincides with compounds which elute as a result of column bleed, which may cause variability (Aggio *et al.*, 2016). This is a plausible explanation as the three earlier eluting compounds (2-pentanone; benzaldehyde; pyridine) in the reference solution were present in all technical replicates. In future, late eluting compounds detected in samples should be interpreted with caution and demonstration of reproducibility of results is required.

4.4.2 Patterns of VOC change

The aim of this chapter is to characterise the faecal VOC profile of mares prior to and following foaling. No episodes of colic occurred in any of the mares during the study period. Esters, aldehydes, acids and ketones made up the majority of chemical classes of compounds

identified in the faeces (list of compounds is in Appendix 4.1). The number of faecal compounds remained relatively stable throughout the study period. Although not significant, it is noteworthy that the fewest compounds were seen at T1 (the week after foaling). A reduction in compound numbers may have occurred because of a reduced level of food intake around the time of foaling. Mare behaviour has been shown to change immediately post-foaling (Van Dierendonck *et al.*, 2004). On the lead up to parturition mares often become uncomfortable and restless, and labour may last several hours (Davies-Morel, 2008). Immediately following parturition, she may be attentive to the foal and extra vigilant during this crucial bonding phase and may have a reduced appetite (Houpt, 2002). It can only be speculated that mare behaviour was responsible for a drop in VOCs around foaling time as behaviour of the mares was not recorded during the study. Mares also experienced a change in management (stabled overnight) at the start of the first week and it is unknown whether this could have an effect on VOCs.

Dimethyl sulfone is formed from dimethyl sulphide along with ketones and is the product (and therefore a possible marker) of amino acid degradation. Digestibility and nutrient content of the feed material were not measured in this work; hence assumptions are made from the literature. Grass is documented to have a higher digestibility than hay (Edouard et al. 2008). Furthermore, crude protein levels are higher in pasture than hay and therefore increased amino acid degradation from grazing is expected (National Research Council, 2007). From T4, apart from some exceptions (Table 4.1) the mares were consuming lush pasture. Dimethyl sulfone increased from T8 onwards (Figure 4.8a - four weeks after the mares had been on lush pasture. The requirements for crude protein and essential amino acids (e.g. lysine) are at their highest during the first month of lactation (National Research Council, 2007). In rats, it has been found that the gut has a greater absorptive capacity for amino acids during lactation than pregnancy (Cripps & Williams, 1975). Dimethyl sulfone however, did not increase immediately after the mares had access to lush pasture. A greater physiological need for amino acids during this time may have resulted in reduced luminal amino acids that may be degraded to dimethyl sulfone (National Research Council, 2007). As with the lowest number of compounds observed at T1, the compound nonane was also at its lowest level of abundance at T1 (Figure 4.8b) and highest during regimen 4, when the mares had access to lush pasture (Figure 4.8b). Two compounds (nonane and 2-octene, 2,6-dimethyl-) showed a trend towards a significant association with time. Nonane is a component of plants, identified in coriander leaf and which disappears after picking which may explain why this compound increased during regimen 4 as the horses were eating grass

rather than conserved forage (Potter & Fagerson, 1990). Nonane has been shown to be produced by bacteria present in the soil and as a metabolite of tall fescue, a common European grass species (Yue *et al.*, 2001; Gu *et al.*, 2007). Previously, 2-octene, 2,6-dimethyl- has been detected in ruminal gases and the authors speculated it to be a product of fermentation (Cai *et al.*, 2006). The change in this compound, over time, may represent subtle changes in fermentation.

Metagenetics (16S rRNA microbiome analysis) was performed on the samples in this work by Salem, (2016). Few differences in OTUs were detected over time or with regimen, indicating a stable microbiota. Stronger clustering was observed for individual mares in the microbiome analysis than the VOC analysis, indicating functional similarity between bacteria. Proudman *et al.* (2014) reported bacteria were stable post-feeding of an oral supplement to enhance pre-caecal digestion. The VOC analysis in Proudman's study showed a more marked change. It was concluded that a metabolic adaptation of existing communities occurred. In the current work, the VOC metabolome was relatively stable with just 1 out of 105 VOCs demonstrating significant change. This could indicate that the management changes mares underwent in the current work were not enough to bring about major shifts in the hindgut bacteria themselves or their function, but further work using a cross-over study design and a larger cohort, for example, is required to confirm this.

Esters were the main drivers of samples clustering in the negative direction of PC1 (Figures 4.5, 4.6 and 4.7). In **Chapter 3** an increase in esters was observed when horses were consuming a grass-based diet. The clustering of T1 and regimen 5 samples away from esters may be because of a lower consumption of grass shortly after foaling and when the mares were stabled for 24 hours. However, significant changes in esters or VFAs were not detected, implying a stable bacterial metabolome. Significant differences in VFAs were not observed in perinatal humans and gut microbiota also remained stable during the perinatal period (Jost *et al.*, 2014). It is worth noting that esters should be interpreted with caution as these were found to increase when stored at -20^{o}C (**Chapter 2**). During the collection period samples were stored at -80^{o}C, only when all samples had been collected was an aliquot transferred to a -20^{o}C freezer for VOC analysis.

4.4.3 Exclusion of T0 and post-anthelmintic and hormone treatment (AH) samples

The samples collected at T0 were stored at 4^{o}C for up to one week before freezing. Others have reported that only 28% of VOCs remained stable after one day of storage at 4^{o}C (human stool) (Phua *et al.*, 2013). It was considered appropriate to check T0 samples did not skew

the results in this study. Storage at 4°C was not investigated in **Chapter 2**, however linear increases in the abundance of some aldehyde compounds were observed with time when samples were exposed to an environment with an average temperature of 6.2°C for 24 hours. Despite this, the removal of T0 and AH samples had very little impact on the overall results (Appendix 4.2 and 4.3). In this work compound numbers or abundance did not appear to increase at T0 (Figures 4.3 and 4.6) and clustering for T0 was not apparent in a PCA plot (Figure 4.5), indicating other factors were stronger than sample storage effect here. The effect of hormone treatments on the equine faecal microbiome and metabolome has not been previously investigated, therefore it was unknown whether these would influence the VOC profile. It has been previously demonstrated that ivermectin can alter faecal microbiome richness post-treatment in yearlings with both low and high FECs (Peachey *et al.*, 2019). However, faecal microbiome richness was not altered post-treatment in broodmares (Peachey *et al.*, 2018). The latter finding agrees with the current work that removal of post-anthelmintic treatment and hormone samples had very little impact on the results in the current study. The effect of the anthelmintics administered to horses in this study (moxidectin and praziquantel) on the microbiome is unknown, but appeared to have very little effect in this work. However, horses had variable worm burdens and further studies to control extrinsic factors and sampling at shorter time-intervals are required.

4.4.3 Comparison to disease

The risk factors for large colon volvulus include: being a broodmare, having foaled in the last 90 days, an increase in hours of stabling within last 14 days, eating hay or sugar beet pulp and experiencing a change in pasture in the last 28 days (Suthers *et al.,* 2013). In the present work, significant VOC changes were associated with time and management regimen. Compared to the number of significant VOCs associated with time seen in **Chapter 3** the patterns observed here were relatively subtle. It is worth noting that the horses in **Chapter 3** also remained healthy. The fact that all mares investigated in the current study remained healthy throughout the collection period may suggest that in this case changes were below a threshold that results in a colic episode. The underlying physiological mechanisms of colic in the broodmare are uncertain. It has been suggested that gut dysbiosis may play some role in colic in the broodmare. Proteobacteria (which include lactic acid producing species) increased in mares prior to them suffering from a colic episode (Weese *et al.* 2014). The healthy mares in the work by Weese *et al.* had minimal change in their faecal bacteria pre- and post-partum. Furthermore, microbiome analysis of samples in the current study also

showed minimal alterations. This evidence indicates that the microbiome and metabolome of the healthy mare remain stable around the time of foaling.

4.4.4 Limitations and future work

Limitations Broodmares have been identified as a group of horses who have an increased risk of colic (Suthers *et al.,* 2013). The samples collected in this work were from mares maintained on a working stud farm, which is assumed to represent a typical management regimen for this category of horse in the UK. The work here observed only subtle changes despite there being alterations in management, administration of drug and hormonal treatments as well as the event of foaling. However, it is debatable that too many confounding factors could have masked the influence of foaling on the faecal VOC profile. A control group of non-pregnant mares undergoing the same management changes but not foaling would have determined whether the subtle alterations in VOC patterns observed were related to foaling or the other factors described. In addition, the sample size of 5 mares was a small representation of the population and may explain the lack of statistical significance observed in this work.

Implications The faecal VOC profile of the healthy periparturient mare 5 weeks before foaling, during second stage labour (T0) and 12 weeks post-foaling has now been established. Current sequencing methods to characterise bacteria are not suitable for everyday diagnosis, whereas portable devices for VOC analysis are becoming cheaper, faster and more readily available than ever before (Agbroko & Covington, 2018). Portable devices would allow the monitoring of faecal VOC changes to identify the onset of a colic episode before it becomes clinically apparent so that preventative action can be taken. This would improve the welfare of mares and foals and reduce the loss of valuable stock.

Future work should involve the sampling of mares before and after a colic episode to determine whether a change in VOC profile precedes a colic episode, as previously found in faecal bacteria (Weese *et al.,* 2014). The present work is the first to study the faecal VOC metabolome within the time-frame that broodmares have been identified to be at the greatest risk of developing colic (Archer & Proudman, 2006). However, as shown in other species, gut microbial changes in pregnancy may begin much earlier (Koren *et al.,* 2012). In order to fully understand the effect of pregnancy on VOCs, future work may involve following mares throughout all trimesters with the aim of being able to predict mares who may develop colic before the high-risk window. Sample collection from multiple farms across the UK would provide a wider representation of the population and a larger sample size.

4.4.5 Overall conclusions

The overall aim of this work is to characterise the temporal faecal VOC changes of the healthy periparturient mare. It was found that the numbers of compounds around this time remained stable along with the majority of fermentation-produced metabolites. The VOC metabolome results reported here mirrored previous work observing a stable faecal microbiome around the time of parturition in the healthy mare. This work provides a benchmark for further investigations into the impact of dietary and management changes on the healthy faecal metabolome of the broodmare as well as making comparisons to disease.

Chapter 5: A comparison of the microbiome and VOC metabolome of colonic contents in *Anoplocephala perfoliata* infected and non-infected horses: an abattoir study.

5.1 Introduction

The gastrointestinal tapeworm *Anoplocephala perfoliata* is a common inhabitant of the equine gut (Pavone *et al.*, 2011; Tomczuk *et al.*, 2014). *A. perfoliata* is of high veterinary importance because of its association with some types of equine colic (Owen *et al.*, 1989; Proudman *et al.*, 1998; Proudman & Holdstockt, 2000; Ryu *et al.*, 2001; Boswinkel & Sloet van Oldruitenborgh-Oosterbann, 2007). Currently licenced anti-cestode drugs (pyrantel and praziquantel) are effective in reducing burdens of *A. perfoliata*. However, there is concern that anthelmintic resistance will become an issue because of the limited drug classes available to treat cestodes in horses (Nielsen, 2016; Lyons *et al.*, 2017). Anthelmintic resistance is a growing concern in other equine parasite species, including the cyathostomins and *Parascaris equorum* (Nielsen *et al.*, 2014). Therefore, alternative methods to diagnose and control equine gastrointestinal parasites are urgently needed.

Intestinal parasites share the gut environment with the intestinal microbiota. Despite this, parasite-microbiota interactions have received very little attention. The involvement of bacteria in parasite establishment and persistence in the gut is a valid area of study; numerous studies reviewed by Zaiss & Harris (2016) have reported more helminths in conventionally raised compared to germ-free mice. Furthermore, antibiotic treatment of helminth-infected mice and pigeons have been shown to reduce worm burdens compared to untreated controls (Hayes *et al.*, 2010; Biswal *et al.*, 2016). Collectively, these studies suggest that the gut microbiota play a role in parasite establishment and maintenance in the gut. Identification of such relationships has consequences for understanding how parasites contribute towards disease of the host (e.g. inflammation and eventually colic in the horse). The development of methods to manipulate the microbiome (without impacting on the host) as an alternative to drugs, to prevent the establishment and persistence of parasites in the gut, would be a very valuable tool. In addition, development of alternative biomarkers to monitor gastrointestinal parasite burdens in the horse utilising faeces would provide benefits to equine health and welfare.

To date, three studies have investigated the correlation between strongyle FECs and the faecal microbiome in the horse (Clark *et al.*, 2018; Peachey *et al.*, 2018, 2019). Clark *et al.,*

(2018) observed modest differences in microbiota composition in ponies identified as resistant or susceptible post-exposure, to natural infection. The work by Peachey *et al.,* (2018) identified differences in faecal microbiome composition between groups of horses with high and low strongyle FECs. Both studies demonstrated an increase in members of the phylum Proteobacteria with high parasite burdens. The main findings of Peachey *et al.,* (2019) were that microbial richness shifted 2 and 14 days post treatment, irrespective of whether FECs were high or low. These results indicated an effect of anthelmintic or clearance of strongyles (even at low levels) from the gastrointestinal tract are factors that alter the equine faecal microbiome. Collectively, these initial studies provide evidence that intestinal helminth-microbiota interactions occur in the horse. Therefore, further work investigating the effect of parasites on the equine gut microbiome is justifiable.

The effect of *A. perfoliata* on the gut microbiome has not previously been investigated. Identification of differences in the gut microbiota between *A. perfoliata* infected and non-infected horses could imply a possible interaction between cestodes and the equine gut microbiota. Furthermore, in the horse the interaction of parasites and the gut metabolome (functional microbiota) has not previously been investigated. As reviewed by Brosschot & Reynolds (2018), studies incorporating both microbiome and metabolome interactions with parasites are important in understanding the functional effect parasites have on the gut microbiota and the host.

The aims of this chapter are to 1) compare the microbiome and VOC metabolome of the colonic contents of *A. perfoliata* infected and non-infected horses and 2) investigate correlations between 16S rRNA sequence data and VOCs.

5.2 Methods

5.2.1 Animals and sample collection

Samples were collected from 51 horses slaughtered for non-experimental purposes at an abattoir in Taunton in November 2015. At post-mortem examination of the caecal lumen, horses were divided into tapeworm infected (TP) and non-tapeworm infected (Co) groups to diagnose *A. perfoliata* by gold standard (the counting of worms present). *A. perfoliata* counts were recorded as high (50+), medium (21-49) and low (1-20). The rationale for choosing these categories is explained in the discussion (section 5.4.1). Luminal contents (~25 ml) were collected from the colon (pelvic flexure) and the rectum. Following collection, samples were transferred on dry ice to the laboratory, divided, and half of each sample was stored at either

-80°C (microbiome analysis) or -20°C (metabolome profiling). Microbiome and metabolome analysis were performed on the colon samples in the current chapter. Metabolome analysis of the rectal samples was performed later, in **Chapter 6**. Additional samples of rectal contents (~10g) were collected into universal containers and transported in a cool box and stored at 4°C upon reaching the laboratory. Faecal egg counts (method detailed in **Chapter 2**) were then performed on the non-frozen samples within 10 days of collection. FECs of 0 - 199 e.p.g were considered as a low, an e.p.g of 200-499 were considered as medium and above 500 e.p.g were recorded as high strongyle. These categories were based on thresholds identified by the American Association of Equine Practitioners (AAEP) and work by Nielsen and colleagues (Kaplan & Nielsen, 2010; Nielsen *et al.*, 2010a, 2013).

5.2.2 Preparation of samples

Portions of colon contents were freeze-dried for 48 hours (Edwards High Vacuum, UK). Two aliquots of freeze-dried faeces were made, one for VOC profiling (100 mg in a glass headspace vial) and one for DNA extraction (100 mg). For DNA extraction two glass beads (4mm, undrilled, Fisher Scientific, UK) were added before the vials were placed in a frozen rack (-80°C) and bead-beaten (TissueLyser II (QIAGEN) for 2 minutes at full power (30 Hz) to break up and mix the fibrous material. The samples were then stored in the freezer (-80°C) until required to complete the next stage of DNA extraction.

5.2.3 DNA extraction

DNA extraction was performed using QIAGEN QIamp DNA stool mini kits (Qiagen Ltd, UK) and steps were followed according to the manufacturer's instructions. As suggested by the manufacturer, after the addition of ASL buffer (Qiagen) to the sample and vortexing it was incubated in a water bath at 95°C for 5 minutes. Heating at this temperature allowed the release of DNA from the lysis of bacterial cells regarded as being difficult to lyse e.g. Gram-positive bacteria. DNA extraction negatives (500 µl ASL buffer instead of sample) were performed to identify any source of contamination from the extraction kit. Subsequently, DNA was quantified using a Qubit (Qubit dsDNA HS assay kit, Life Technologies) and the sample diluted to 1ng DNA/ µl. Aliquots of DNA were then stored at -80°C until required for the next stage.

5.2.4 Amplicon specific polymerase chain reaction (PCR)

To amplify the bacterial 16S rRNA gene the primers used were F515 ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT NNN NNG TGC CAG CMG CCG CGG TAA and R806 GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T GG ACT ACH VGG GTW TCT AAT (HPLC grade, IDT) (Bates *et al.*, 2011). For the same as **Chapter 3**, the red part of the primer sequence is overhang and complementary to the barcoded set of primers used for the second round of PCR. A PCR master mix was made containing (per 20 µl reaction): Q5 reaction buffer 1x, 0.02 U/µl Q5 High-Fidelity DNA polymerase, 200 µM dNTPs (NEB), 0.125 µM of each forward and reverse primers made up to 15 µl with molecular grade water. The master mix was divided between wells, with 5 µl (1 ng/ µl) of template DNA. The plate was loaded into a pre-heated thermocycler (Multigene^{TM} and Multigene^{TM} OptiMax, Labnet International) with an initial denaturation of 2 minutes at 98^{o}C followed by 14 cycles of 20 seconds at 95^{o}C, 15 seconds at 65^{o}C, 30 seconds at 72^{o}C. There was a final extension step of 5 minutes at 72^{o}C and then samples were held at 4^{o}C. The first round of PCR was carried out in duplicate to reduce PCR bias. The PCR product was then stored at 4^{o}C until purification.

5.2.5 Purification with Axygen beads

The purification step was carried out as detailed in **Chapter 3**.

5.2.6 Index polymerase chain reaction (PCR)

Index PCR was performed as described in **Chapter 3**. The set of barcoded index primers used are listed in Appendix 5.1. The final elute was 25 µl and the concentration of DNA was recorded using a Qubit (Qubit dsDNA HS assay kit, Life Technologies).

5.2.7 Sample pooling and MiSeq Illumina next generation sequencing

Sample pooling and quality control steps were carried out by the CGR as detailed in **Chapter 3**. Paired end (2x250 bp) sequencing was performed on an Illumina MiSeq platform in the Centre for Genomic Research (GCR) facility, University of Liverpool.

5.2.8 VOC profiling

Aliquots of non-freeze-dried and freeze-dried material were analysed by GCMS as detailed in **Chapter 2**. The SPME fibre used was a DVB-CAR-PDMS. Analysis of non-freeze-dried material was performed in triplicate (n=3). A smaller volume of sample was available for the freeze-dried material which was analysed without replicates.

5.2.9 Data processing of 16S rRNA and VOC data

Illumina adapter sequences were trimmed from raw FastQ files using Cutadapt (1.2.1) (Martin, 2011). The reads were further trimmed using Sickle (1.2) (Joshi & Fass, 2011) with a minimum window quality score of 20. Reads shorter than 200 bp after trimming were removed. Samples with fewer than 1000 reads were also removed at this stage. Sequence clustering (99% similarity) was performed by SWARM 2.0 (d=3) (Mahé *et al.*, 2015) and chimeras were filtered out using UCHIME (Edgar *et al.*, 2011). The following steps were then carried out using Quantitative Insights into Microbial Ecology (QIIME) (1.9.1) (Caporaso *et al.*, 2010b). The classifier tool BLAST (Edgar, 2010) was applied together with the Greengenes (gg_13_8) (McDonald *et al.*, 2012) database to assign OTUs at a 99% threshold. The final steps were alignment of OTUs with the database using PyNAST (Caporaso *et al.*, 2010a) and construction of a phylogenetic tree using FastTree (Price *et al.*, 2010). Dr Alessandra Frau (University of Liverpool) performed the data processing steps of the 16S rRNA data.

Data processing and identification of VOCs was performed as described in **Chapter 2.**

5.2.10 Statistical analysis

For both 16S rRNA sequence data and VOC data, statistical analysis was performed on two subsets of the data – all tapeworm (AT) vs control (CO) and those with $\geq$ 21 tapeworms, defined as medium and high (MH) vs CO. The rationale for this analysis is explained in the discussion (section 5.4.1). Statistical analysis was not performed to compare controls with low, medium or high tapeworm groups separately because of unbalanced sample sizes. However, visual comparisons for 16S rRNA sequence data (taxonomy plots), PCA and box plots of compound abundance were plotted for VOC data to show change with level of tapeworm burden.

Appropriate statistical analysis of 16S rRNA (diversity indices, compositional differences) and VOC data (PCA and differences in VOC abundance) was also performed on a sub-group of samples assigned to either FEC of $\geq$200 e.p.g (HSC, n=12) or FEC of $\leq$ 10 e.p.g (LSC, n=11), regardless of tapeworm infection status, to compare with the findings of Peachey *et al.,* (2018).

5.2.10.1 Microbiome statistical analysis

For visualisation purposes, taxonomy plots were constructed for phylum, class, order and family. The data were normalised to relative abundance for these visualisation plots.

Diversity indices were calculated at OTU level using the R package vegan. The alpha diversity (Richness, Shannon and Fisher indices) was calculated for the CO and tapeworm groups and pair-wise differences were compared using the aov() function in R. The beta diversity was plotted as non-metric multidimensional scaling (NMDS) ordination plots (distances: Bray, Unifrac and Weighted Unifrac). Statistical difference in the beta diversity between groups was assessed by PERMANOVA using the adonis() function.

Differential abundance of taxa between CO and tapeworm groups was evaluated using the DESeq2 R package (Love *et al.*, 2014). The DESeq2 package was originally developed for RNA-seq analysis to identify differentially expressed genes and has recently been applied to metagenetics studies (Penington *et al.*, 2018). Data were not normalised before analysis using the DESeq2 package, as the package performs its own normalisation steps, including for adjusting to library size (Love *et al.*, 2014). Differential abundance was carried out at phylum, order, class, family, genus and OTU level. Dr Umer Zeeshan Ijaz (University of Glasgow) wrote the microbiome analysis scripts.

5.2.10.2 VOC metabolome statistical analysis

VOC numbers detected in freeze-dried and non-freeze-dried colon contents were compared using a Wilcox signed rank test with continuity correction, as the data were not normally distributed (Shapiro-Wilk test). Greater numbers of VOCs were detected in the freeze-dried material (Results section 5.3.4.1); therefore, these samples were taken forward for statistical analysis.

The number of VOCs in tapeworm and CO samples were compared using a t-test. The presence or absence of VOCs in groups were compared using Fisher's exact test. Half-minimum values were then imputed for 'missing data' to allow statistical analysis (PERMANOVA and multivariate multiple regression) to be performed on VOC abundance. PCA of VOC abundance was used to check for clustering according to GCMS batch, tapeworm burden and strongyle FEC (categorised as low or medium). Strongyle FEC values rather than categories were also included in the statistical models to further assess the effect of this variable. PCA, PERMANOVA and a t-test comparing VOC abundance were performed to compare HSC and LSC samples.

5.2.11 Identification and removal of outliers for both datasets

Three outliers (C11, C25 and T6) were identified by visual inspection of the taxonomy analysis (Appendix 5.2) of the 16S rRNA data and PCA analysis of VOC data. Samples T6 and C11

contained higher relative abundances of Fusobacteria and Proteobacteria, respectively, compared to Bacteroidetes. High levels of Proteobacteria or Fusobacteria may be indicative of gut dysbiosis (Costa *et al.*, 2012). Therefore, because the history of the animals in this work was unknown, these samples were removed prior to statistical analysis. A third outlier (C25) identified in the VOC data only (Appendix 5.2) was removed as it failed to freeze-dry completely.

5.2.12 Integration of microbiome and metabolome data

The R package mixOmics using the DIABLO framework was applied to microbiome and VOC data using the same methods described in **Chapter 3**. The model was performed to compare the AT (n=21) and CO (n=48) samples.

5.3 Results

5.3.1 Parasitology

A total of 21 (41%) horses had *A. perfoliata* present. Horses with low, medium and high burdens are recorded in Table 5.1. The number of horses without *A. perfoliata* present was 59%. Strongyle (eggs were ovoid, white and thin-walled) FEC values for each horse are also shown in Table 5.1. Samples were also checked for ascaris eggs (rounded and brown with a thick outer-pitted wall) and all were negative.

Horse ID (Control)	**FEC (e.p.g)**	**Horse ID (Tapeworm positive)**	***A. perfoliata* status**	**FEC (e.p.g)**
C1	10	T1	low	90
C2	423	T2	unrecorded	19
C3	141	T3	unrecorded	0
C4	42	T4	low	12
C5	156	T5	low	273
C6	180	T6	med	216
C7	0	T7	med	22
C8	72	T8	low	300
C9	234	T9	unrecorded	177
C10	135	T10	unrecorded	105
C11	3	T11	unrecorded	162
C12	0	T12	high	43
C13	23	T13	high	48
C14	174	T14	med	315
C15	135	T15	high	54
C16	288	T16	high	31
C17	159	T17	med	66
C18	219	T18	high	0
C19	252	T19	med	135
C20	174	T20	low	2
C21	0	T21	low	1
C22	120			
C23	0			
C24	204			
C25	315			
C26	0			
C27	13			
C28	216			
C29	99			
C30	282			

Table 5.1 Parasitology results of 51 horses. For both controls and tapeworm positive horses strongyle faecal egg count (FEC) was recorded. For the tapeworm infected horses, the level of infection was recorded as low (1-20 worms), medium (21-49 worms) or high (50+ worms).

5.3.2 Microbiome analysis results

The total number of sequences obtained post filtering was 6,186,806. The sequences clustered to a total number of 1,458. There was a minimum of 75,244 sequences per sample and a maximum of 167,089, averaging at 116,212 sequences per sample.

5.3.2.1 Taxonomic summary

Visual taxonomic summaries for phylum, class, order and family for all samples (excluding outliers) are shown in Figures 5.1 and 5.2. The most dominant phyla were Firmicutes (51.6%), Bacteroides (36.1%), Spirochaetae (2.3%) and Fibrobacteres (2.0%).

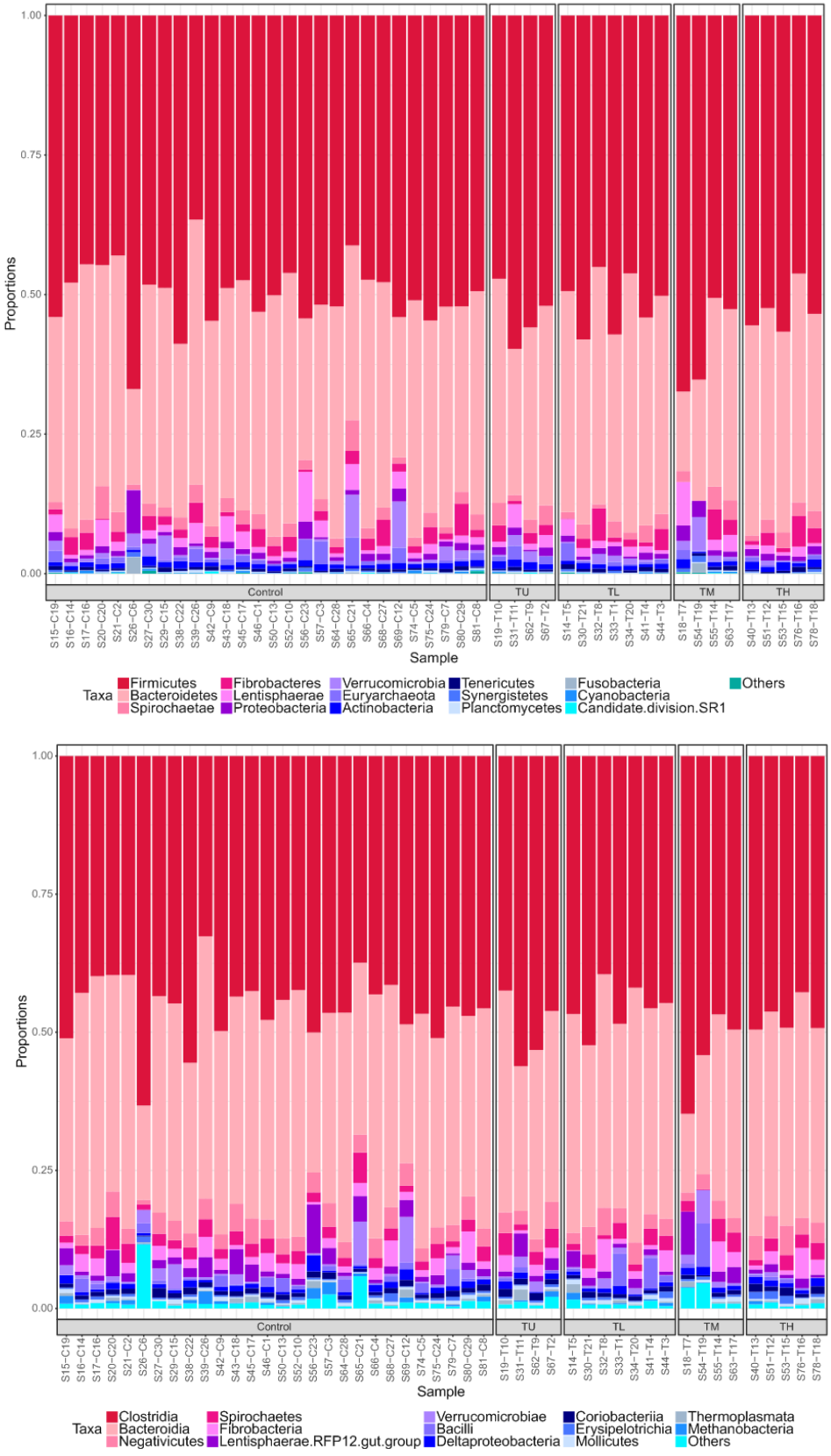


Figure 5.1 Taxonomic summaries for a) phylum, b) class (no outliers) of the colonic contents of 48 horses infected and not infected with *A. perfoliata*. Key: TU = tapeworm positive with burden unrecorded, TL = low tapeworm, TM = medium tapeworm, TH = high tapeworm

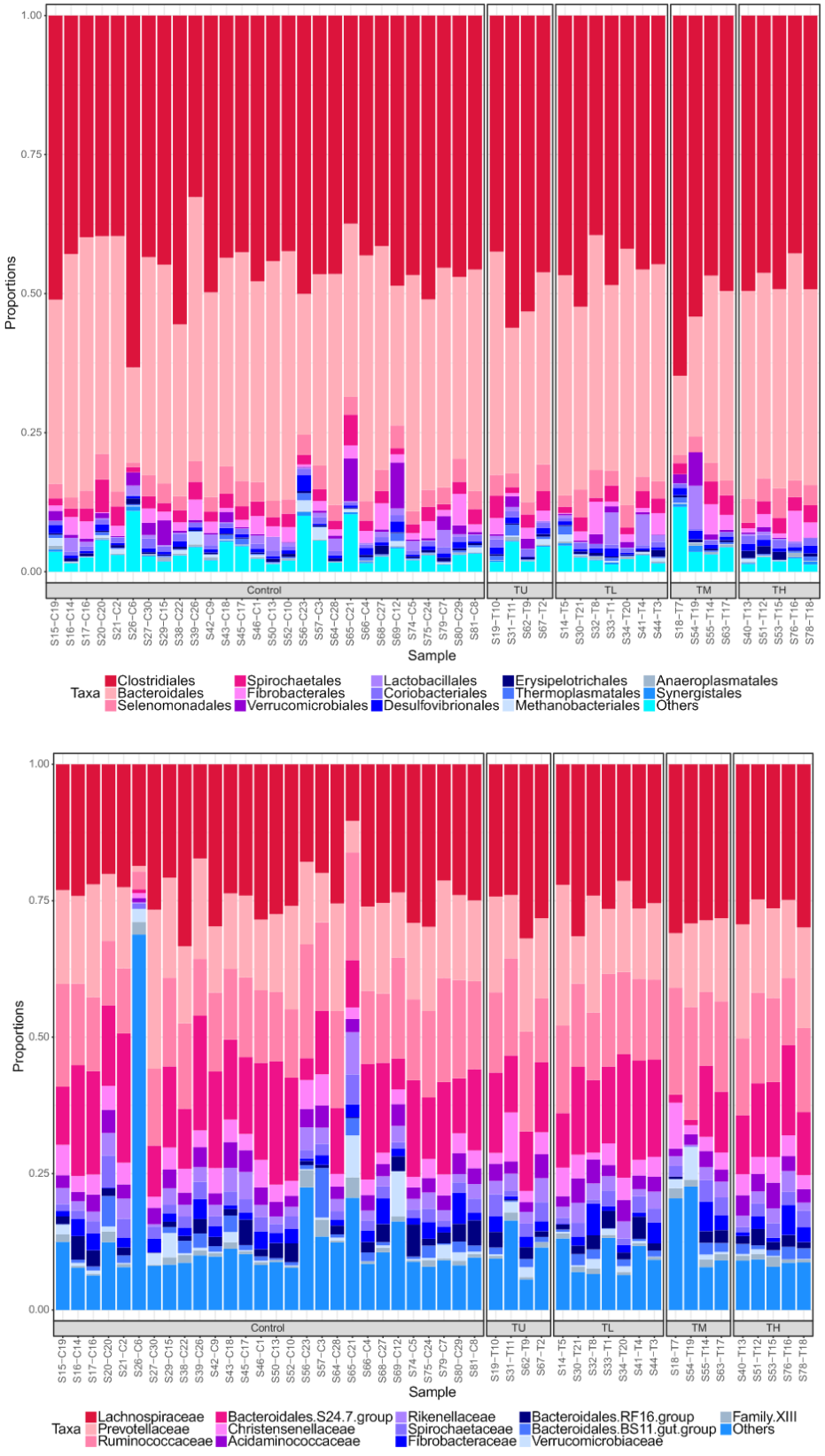


Figure 5.2 Taxonomic summaries for a) order, b) family (no outliers) of the colonic contents of 48 horses infected and not infected with *A. perfoliata*. Key: TU = tapeworm positive with burden unrecorded, TL = low tapeworm, TM = medium tapeworm, TH = high tapeworm

5.3.2.2 Diversity indices

Box plots of the alpha diversity indices (Richness, Shannon and Fisher alpha) are shown in Figure 5.3 a). (AT vs Co) and b). (MH vs CO). Pair-wise comparisons between groups were not significant (p>0.05). Values for beta diversity are shown in Table 5.2 for both AT and MH groups compared to controls. PERMANOVA of beta diversity values revealed there were no significant differences in beta diversity between groups. NMDS plots of the beta diversity (bray) are shown in Figure 5.3 a). (AT vs CO) and b). (MH vs CO) and demonstrated no distinct clustering of sample groups.

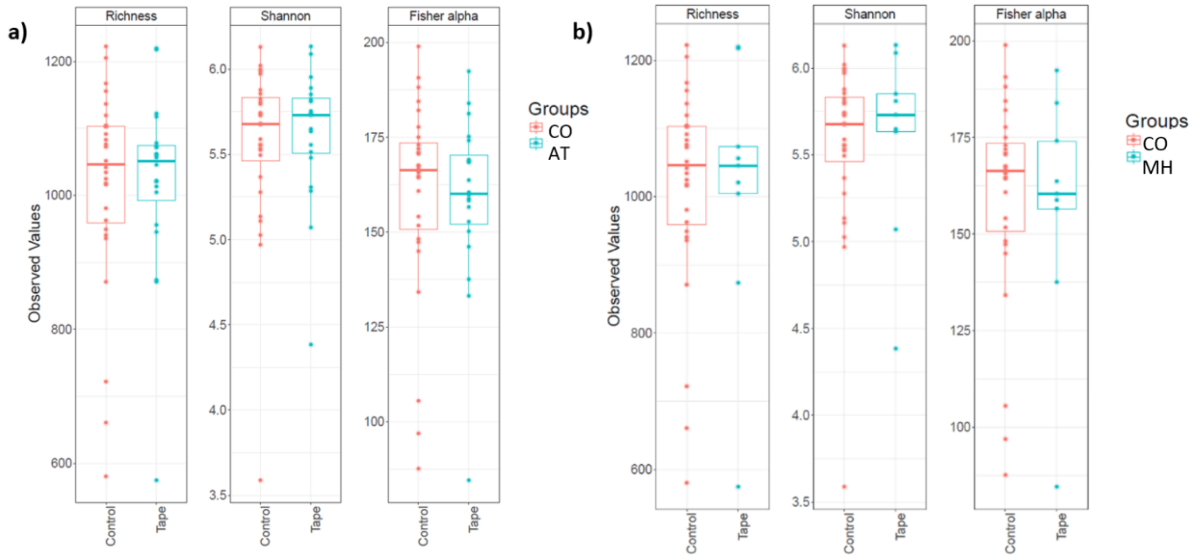


Figure 5.3 Box plots of the bacterial alpha diversity of the colonic contents of 48 horses infected and not infected with *A. perfoliata*. Alpha diversity was not significantly different between groups, pair-wise ANOVA (p>0.05). **a** CO (n=28) vs AT (n=20), **b** CO vs MH (n=9). Key: AT = all tapeworm samples, MH = tapeworm samples with 21+ worms, CO = control (tapeworm negative).

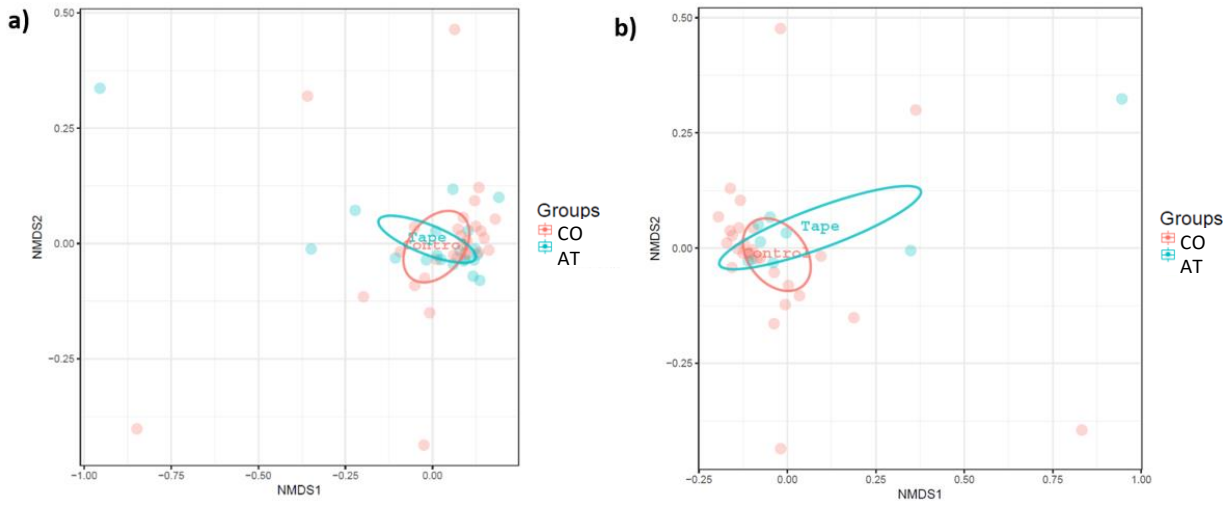


Figure 5.4 Beta diversity NMDS plots (Bray distance is shown) of the colonic contents of 48 horses infected and not infected with *A. perfoliata.* a) shows AT (n=20) and CO (n=28), **b)** CO and MH (n=9). The beta diversity between groups was not significant as shown in Table 5.2 (PERMANOVA, p>0.05). Key: AT = all tapeworm samples, MH = tapeworm samples with 21+ worms, CO = control (tapeworm negative).

Comparison	Bray	Weighted Unifrac	Unifrac
AT vs CO	R^2 = 0.03, p = 0.14	R^2 = 0.03, p = 0.25	R^2 = 0.03, p = 0.14.
MH vs CO	R^2 = 0.04, p=0.12	R^2 = 0.04, p=0.07	R^2 = 0.04, p = 0.13

Table 5.2 PERMANOVA results comparing the beta diversity indices between tapeworm-infected groups and the control. Key: AT = all tapeworm samples (n=20), MH = tapeworm samples with 21+ worms (n=9), CO = control (tapeworm negative) (n=28).

5.3.2.4 Differential analysis

When CO and AT were compared, there were no significant differences in taxa abundance at phylum, order or class levels. Some differences at family and genus level were observed and are listed in Table 5.3. At OTU level, 69 OTUs (listed in Appendix 5.3) were found to be significantly different; of these 17 (24.6%) were more abundant in the AT group. When CO and MH were compared, significant differences in taxa between groups were observed at order, family, genus (Table 5.3) and OTU level. One hundred and eighteen OTUs (shown in Appendix 5.4) were significantly different with 28 (23.7%) more abundant in the MH group.

Taxon	p-value (FDR corrected)	Higher abundance in
CO vs AT		
Family		
Bacteroidales UCG-001	0.016	CO
Genus		
Ruminococcaceae UCG-004	0.006	CO
Jeotgalicoccus	0.012	AT
Candidatus Soleaferrea	0.012	CO
Romboutsia	0.028	CO
CO vs MH	**p-value (FDR corrected)**	**Higher abundance in**
Order		
Rickettsiales	0.004	CO
Family		
Rickettsiales Incertae Sedis	0.005	CO
Genus		
Candidatus Hepatincola	0.01	CO
Selenomonas 3	0.02	MH

Table 5.3 Taxa that were significantly different in abundance between tapeworm groups and controls. Differential abundance was calculated by the DESeq2 R package. Key: AT = all tapeworm samples (n=20), MH = tapeworm samples with 21+ worms (n=9), CO = control (tapeworm negative) (n=28).

5.3.2.5 Comparison of the microbiome of samples with a high and low FEC

There were no significant differences in alpha (Richness, Shannon and Fisher alpha) or beta diversity (Bray, Unifrac and Weighted Unifrac). Alpha and beta diversity plots are shown in Figure 5.5a and b Differential analysis revealed 93 significantly different OTUs; 59.1% were more abundant in HSC and 40.9% were more abundant in LSC. A list of taxa significantly different in abundance between LSC and HSC are shown in Table 5.4. OTUs identified as significantly different in abundance between LSC and HSC are shown in Appendix 5.5.

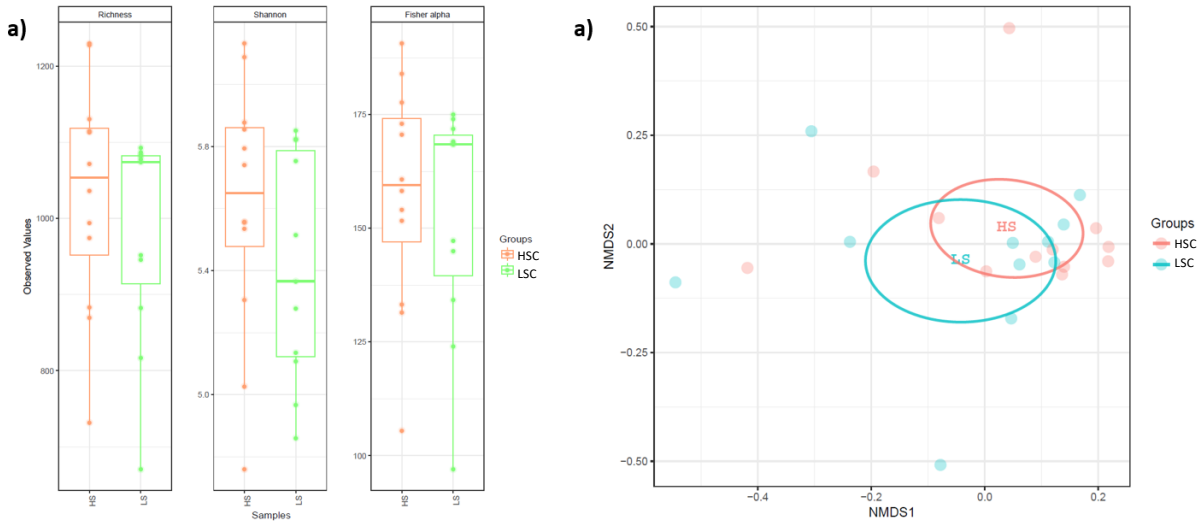


Figure 5.5 Alpha (a) and beta (b) (Bray distance is shown) diversity indices for horses with low (≤ 10 e.p.g) and high (≥ 200 e.p.g) strongyles FECs. Key: HSC = high strongyles (≥ 200 e.p.g), LSC = low strongyles (≤ 10 e.p.g), FEC = faecal egg count.

Taxon	p-value (FDR corrected)	Higher abundance in
HSC vs LSC		
Order		
Rhodospirillales	0.004	HSC
Pasteurellales	0.015	LSC
Family		
Rhodospirillaceae	0.006	HSC
Genus		
Anaerosporobacter	0.001	HSC
Thalassospira	0.006	HSC
Prevotellaceae Ga6A1 group	0.006	HSC
Ruminococcaceae UCG-009	0.013	HSC
Actinobacillus	0.013	LSC
Haemophilus	0.027	LSC
Gordonibacter	0.045	HSC
Cellulosilyticum	0.045	LSC

Table 5.4 Taxa which were significantly different in abundance LSC and HSC groups. Differential abundance was calculated by the DESeq2 R package. Key: LSC = low strongyles ($\leq$ 10 e.p.g), HSC = high strongyles ($\geq$ 200 e.p.g).

5.3.4 VOC metabolome results

5.3.4.1 Comparison of VOCs from freeze-dried and non-freeze-dried colon contents

A significantly greater number of compounds were detected in freeze-dried rather than non-freeze-dried samples of colon contents (Wilcoxon signed-rank test with continuity correction, $p<0.001$). A boxplot demonstrating the spread of the data is shown in Figure 5.6. A chromatogram overlay of the freeze-dried and non-freeze-dried material from one horse (T4) is shown in Figure 5.7.

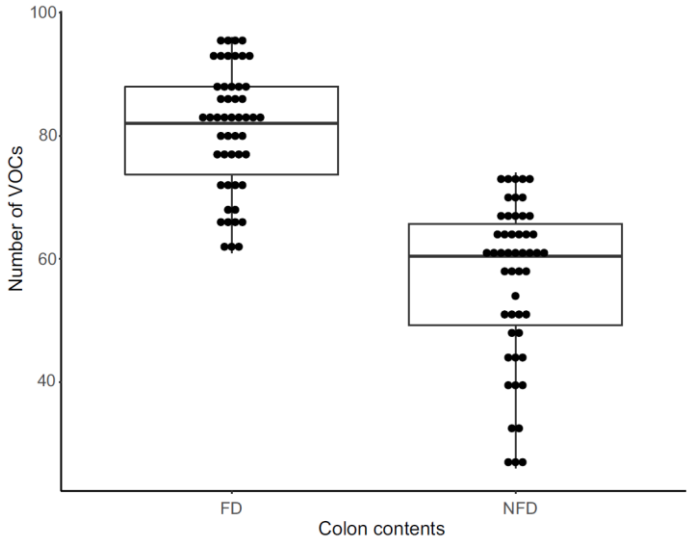


Figure 5.6 Boxplot of the number of VOCs detected in non-freeze-dried (NFD) and freeze-dried (FD) equine colon contents. The weight of the samples was 1000 mg for NFD and 100 mg for FD (after freeze-drying).

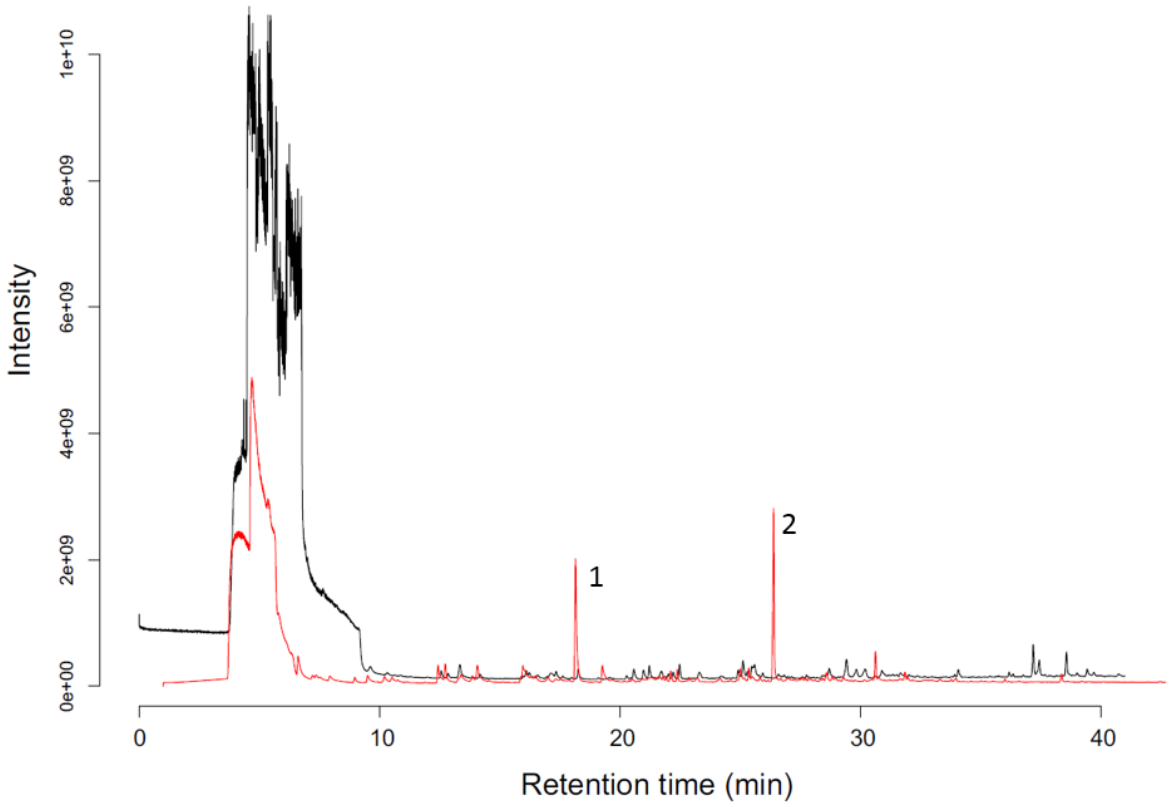


Figure 5.7 A GCMS chromatogram overlay of the VOC profile of non-freeze-dried (black trace) and freeze-dried (red trace) colon contents from one horse (ID: T4). Peak 1 = hexanal, peak 2 = dimethyl sulfone.

5.3.4.2 Numbers and presence of VOCs in tapeworm and control samples

In total 84 VOCs were identified across all 51 samples. A complete list with the VOCs identified within TP and CO samples is shown in Appendix 5.6. A Shapiro-Wilk test revealed that VOC numbers were normally distributed (p=0.43). A significantly higher mean number of VOCs were detected in the AT group (mean 72 SE $\pm$ 1.44) compared to CO (67 $\pm$ 1.21) (p=0.03, t-test). There were no VOCs that significantly differed in absence or presence in samples between the AT and CO groups (p>0.05, Fisher's exact test). For the MH group the mean number of VOCs (67 $\pm$ 2.43) was not significantly different from CO (72 $\pm$ 1.21, p=0.12, t-test). There were no VOCs that significantly differed in absence or presence in samples between the MH and CO (p>0.05, Fisher's exact test). There were no significant differences in numbers or absence or presence of VOCs between horses with high and low strongyle FEC (p>0.05, t-test and Fisher's exact test, respectively).

5.3.4.3 PCA plots

A PCA plot for all samples labelled by GCMS batch is shown in Figure 5.8: there was no apparent clustering for batches of samples run on different days. A PCA including AT and CO samples is shown in Figure 5.8a, and a PCA for MH and CO samples is shown in Figure 5.9b. Clustering was not evident for the AT vs CO groups, but 6/9 MH samples clustered in the positive direction of PC1 (Figure 5.9b). The 10 VOCs with the highest loading scores (0.15-0.19) in the positive direction of PC1 were 5-hepten-2-one, 6-methyl-; furan, 2-pentyl-; 2-heptanone; 2-octanone; furan, 2-ethyl-; cyclohexanone, 2,2,6-trimethyl-; 1-cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-; 2(3H)-furanone 5-ethyldihydro-; 1-octen-3-ol and nonane. No apparent clustering was evident for Strongyle FEC when split into either a medium and low category or into LSC and HSC (Figure 5.10a and b).

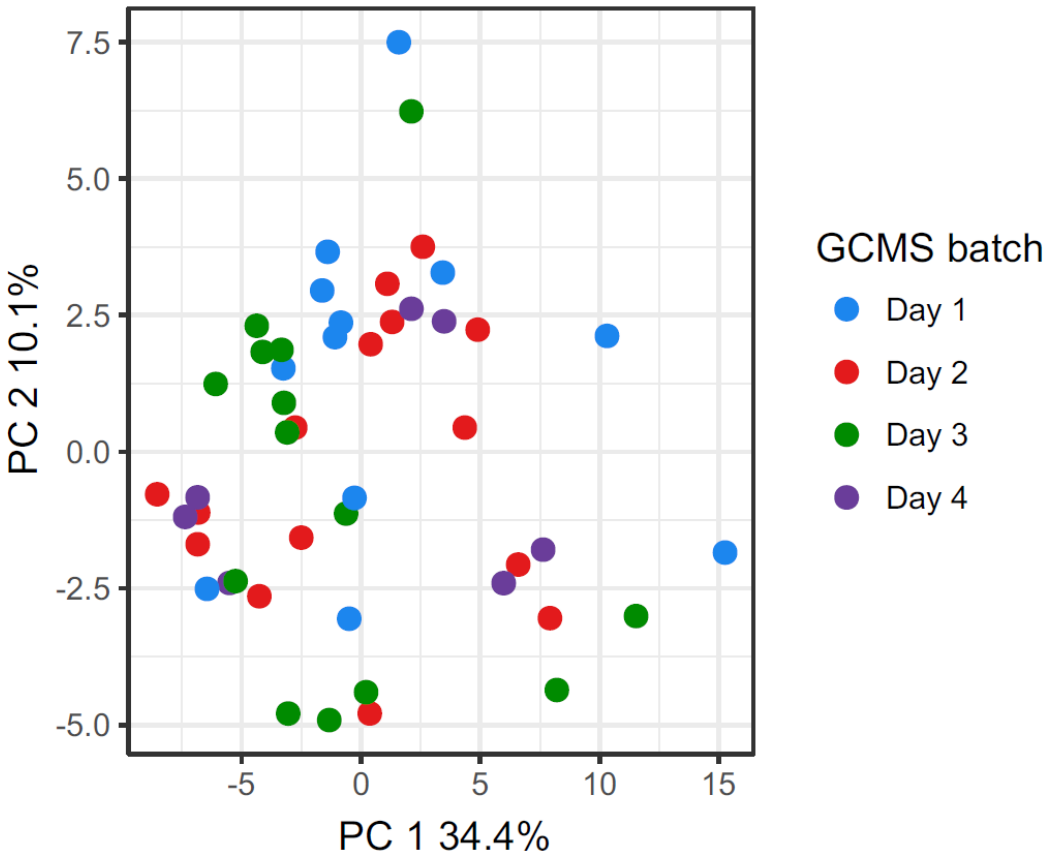


Figure 5.8 A PCA plot of the VOC profiles of the colonic contents of 48 horses infected and not infected with *A. perfoliata*. Samples labelled for GCMS batch (days 1-4).

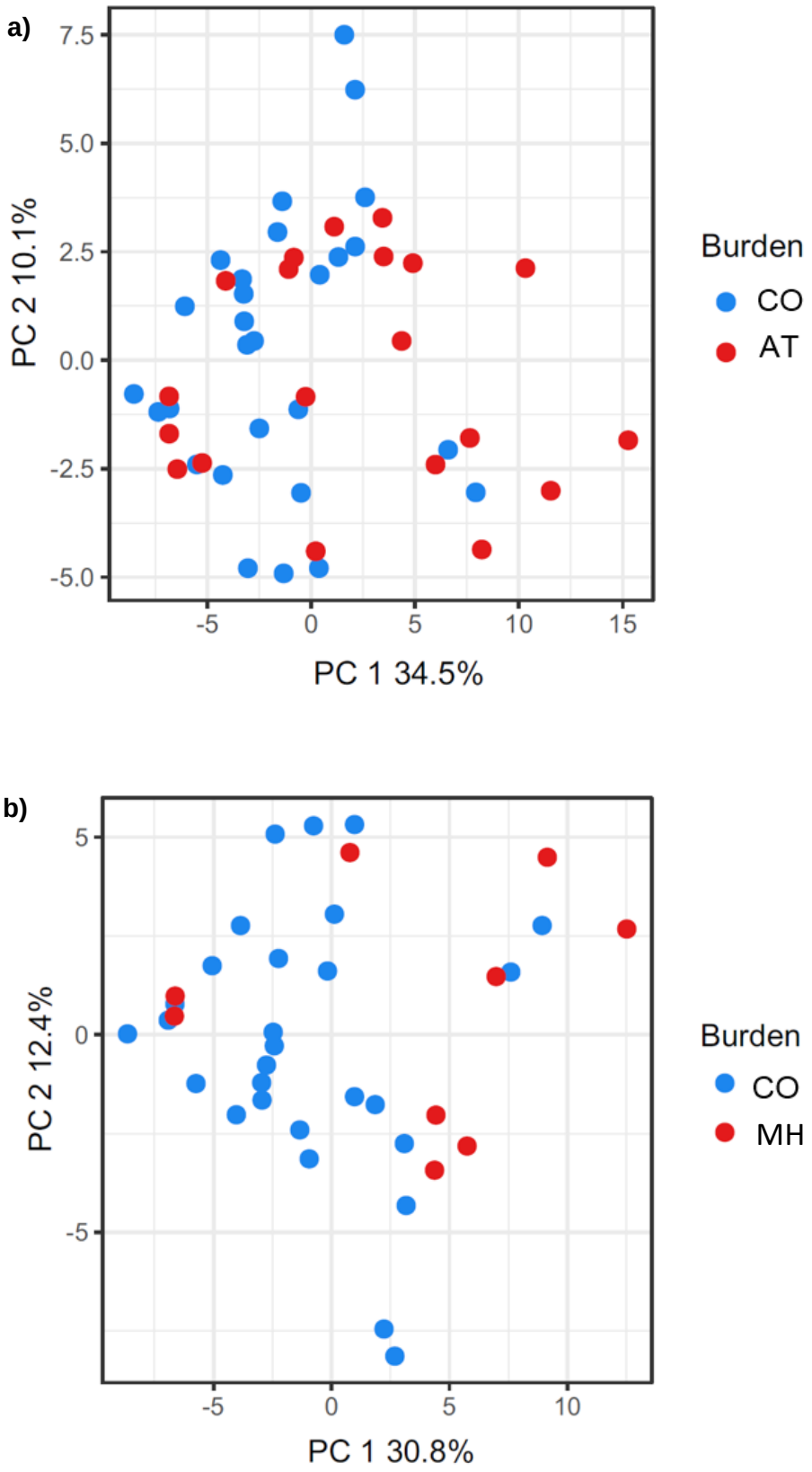


Figure 5.9 PCA plots of the VOC profiles of the colonic contents of 48 horses infected and not infected with *A. perfoliata*. Horses with 1+ worms (AT) and control (CO) are shown in **a)** Horses with 21+ worms (MH) and CO are shown in **b)**.

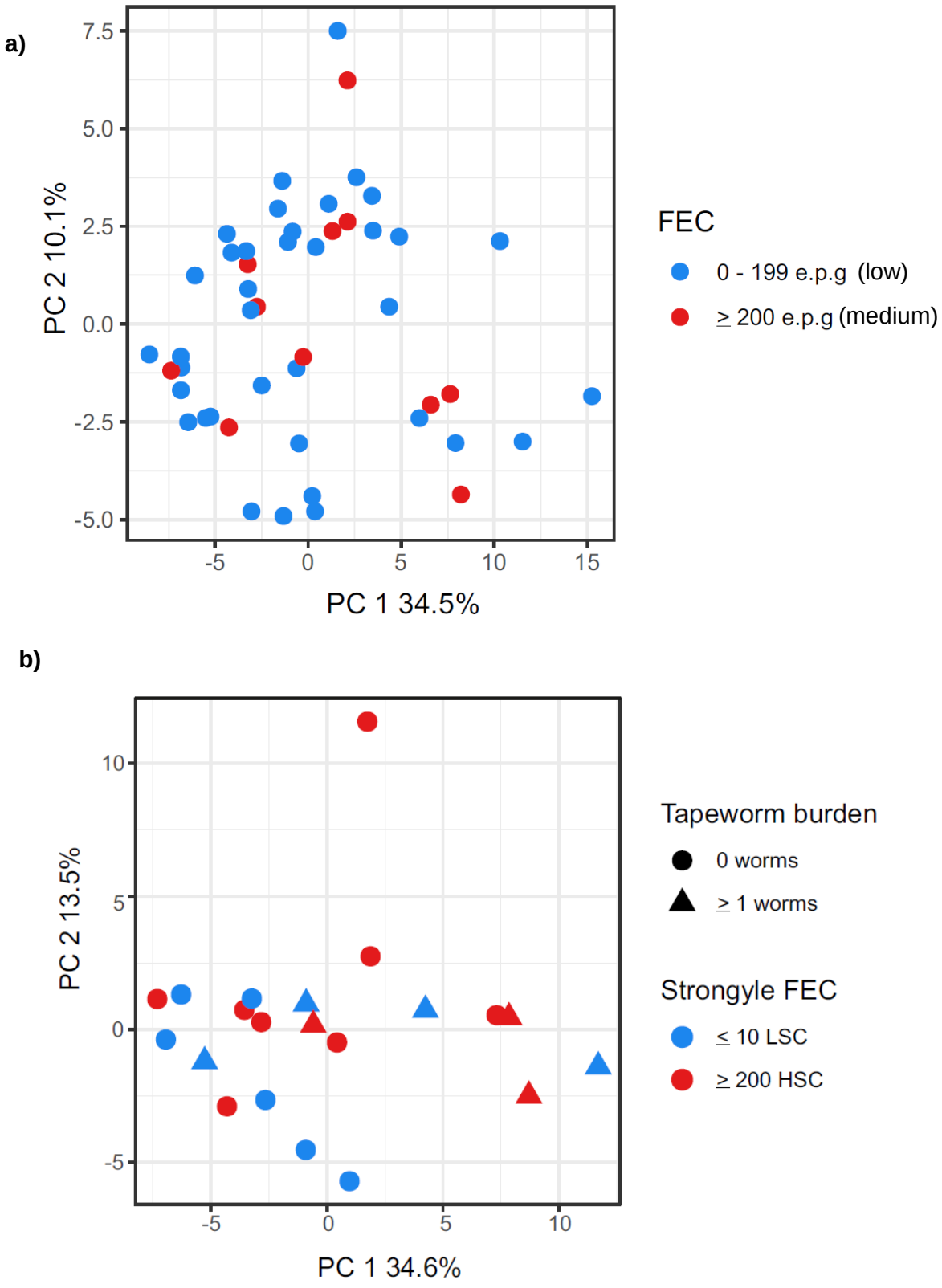


Figure 5.10 PCA plots of the VOC profiles of the colonic contents horses labelled for strongyle FEC. Horses are grouped for low and medium FEC in **a)**. In **b)** horses are grouped into ≤ 10 e.p.g or ≤ 200 e.p.g categories. The symbol also represents whether the horses were positive or negative for *A. perfoliata.* HSC = high strongyles (≥ 200 e.p.g), LSC = low strongyles (≤ 10 e.p.g), FEC = faecal egg count.

5.3.4.4 The association of tapeworm burden and VOC abundance

Twenty-six VOCs were significantly associated with tapeworm burden and three associated with Strongyle FEC when AT and CO were included in a multivariate multiple regression model (pre-correction for multiple comparisons). For MH and CO, 23 VOCs were associated with tapeworm burden and one was associated with Strongyle FEC. However, after correction for multiple comparisons, none of the VOCs was significantly different in abundance between horses with tapeworm and controls for both the AT and MH groups. Table 5.5 and Table 5.6 list VOCs which were significant before Benjamin-Hochberg correction and the adjusted p-values for AT vs CO and MH vs CO, respectively. The top three VOCs with the lowest p-values for both comparisons are plotted in Figure 5.11. Box plots were constructed to show the linear relationship of tapeworm burden and VOC abundance for some VOCs. The results of a PERMANOVA analysis to assess how much variation could be described in the VOC profile by tapeworm burden and strongyle FEC is shown in Table 5.7.

AT group	P-value	Adjusted p-value	Fold change	Higher abundance in
VOCs associated with tapeworm burden				
Dodecane, 2,6,10-trimethyl-	0.006	0.088	-0.57	AT
Undecane, 2,6-dimethyl-	0.006	0.088	-0.60	AT
Acetic acid	0.009	0.088	-0.98	AT
Nonane	0.009	0.088	-0.61	AT
2-Decanone	0.009	0.088	-1.07	AT
2-Nonanone	0.010	0.088	-1.38	AT
2-Undecanone	0.011	0.088	-0.99	AT
Furan, 2-methyl-	0.012	0.088	-1.25	AT
1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	0.013	0.088	-1.70	AT
2-Butanone, 4-(2,6,6-trimethyl-1-cyclohexen-1-yl)-	0.013	0.088	-0.67	AT
Heptane, 2-methyl-	0.014	0.088	-0.66	AT
1-Octen-3-ol	0.014	0.088	-0.67	AT
2-Heptanone, 6-methyl-	0.016	0.088	-0.94	AT
2-Octanone	0.018	0.088	-1.15	AT
3-Octanone	0.018	0.088	-0.80	AT
3-Pentanone, 2-methyl-	0.019	0.088	-1.40	AT
Propanal	0.021	0.088	-0.47	AT
1-Penten-3-one	0.021	0.088	-0.65	AT
2-Heptanone	0.021	0.088	-0.91	AT
1-Heptanol	0.021	0.088	0.33	CO
2-Octene, (E)-	0.022	0.088	-0.57	AT
Furan, 2-pentyl-	0.024	0.092	-1.38	AT
D-limonene	0.026	0.097	-0.57	AT
Cyclohexanone, 2,2,6-trimethyl-	0.027	0.097	-1.03	AT
Methacrolein	0.034	0.115	-0.55	AT
1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.040	0.132	-1.12	AT
VOCs associated with FEC	0.014	0.600	NA	NA
Heptane, 2-methyl-	0.020	0.600	NA	NA
1-Octen-3-ol	0.027	0.600	NA	NA
2-Butanone, 4-(2,6,6-trimethyl-1-cyclohexen-1-yl)-	0.014	0.600	NA	NA

Table 5.5 Multivariate multiple regression modelling of VOC abundance of the colonic contents of horses infected with *A. perfoliata* (1+ worms) and non-infected controls. Fold-change between AT and CO is also shown. Fold-change could not be calculated for FEC as values were not divided into categories for the model. Key: AT = all tapeworm samples, CO = control (tapeworm negative), FEC = faecal egg count.

MH group	P-value	Adjusted p-value	Fold change	Higher abundance in
VOCs associated with tapeworm burden				
Undecane, 2,6-dimethyl-	0.003	0.138	-0.70	MH
Heptane, 2-methyl-	0.005	0.138	-0.88	MH
2-Heptanone, 6-methyl-	0.006	0.138	-1.09	MH
1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	0.006	0.138	-1.90	MH
2-Decanone	0.012	0.146	-0.95	MH
2-Undecanone	0.014	0.146	-1.24	MH
Furan, 2-pentyl-	0.016	0.146	-1.56	MH
Octane	0.017	0.146	-0.61	MH
1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.017	0.146	-1.40	MH
1-Octen-3-ol	0.018	0.146	-0.74	MH
2-Octene, (E)-	0.019	0.146	-0.66	MH
Cyclohexanone, 2,2,6-trimethyl-	0.021	0.146	-1.23	MH
Ethyl acetate	0.027	0.154	0.79	CO
Dodecane, 2,6,10-trimethyl-	0.030	0.154	-0.56	MH
2-Nonanone	0.031	0.154	-1.27	MH
Decane	0.033	0.154	-0.66	MH
Nonane	0.033	0.154	-0.65	MH
Nonane, 2-methyl-	0.033	0.154	-0.74	MH
Acetic acid	0.037	0.154	-0.92	MH
2-Octanone	0.038	0.154	-1.04	MH
5-Hepten-2-one, 6-methyl-	0.039	0.154	-1.31	MH
2-Heptanone	0.040	0.154	-0.95	MH
2-Butanone, 4-(2,6,6-trimethyl-1-cyclohexen-1-yl)-	0.043	0.158	-0.72	MH
VOCs associated with FEC				
Butanal, 3-methyl-	0.026	0.991	NA	NA

Table 5.6 Multivariate multiple regression modelling of VOC abundance of the colonic contents of horses infected with *A. perfoliata* (21+ worms) and non-infected controls. Fold-change between MH and CO is also shown. Fold-change could not be calculated for FEC as values were not divided into categories for the model. Key: MH = tapeworm samples with 21+ worms, CO = control (tapeworm negative), FEC = faecal egg count.

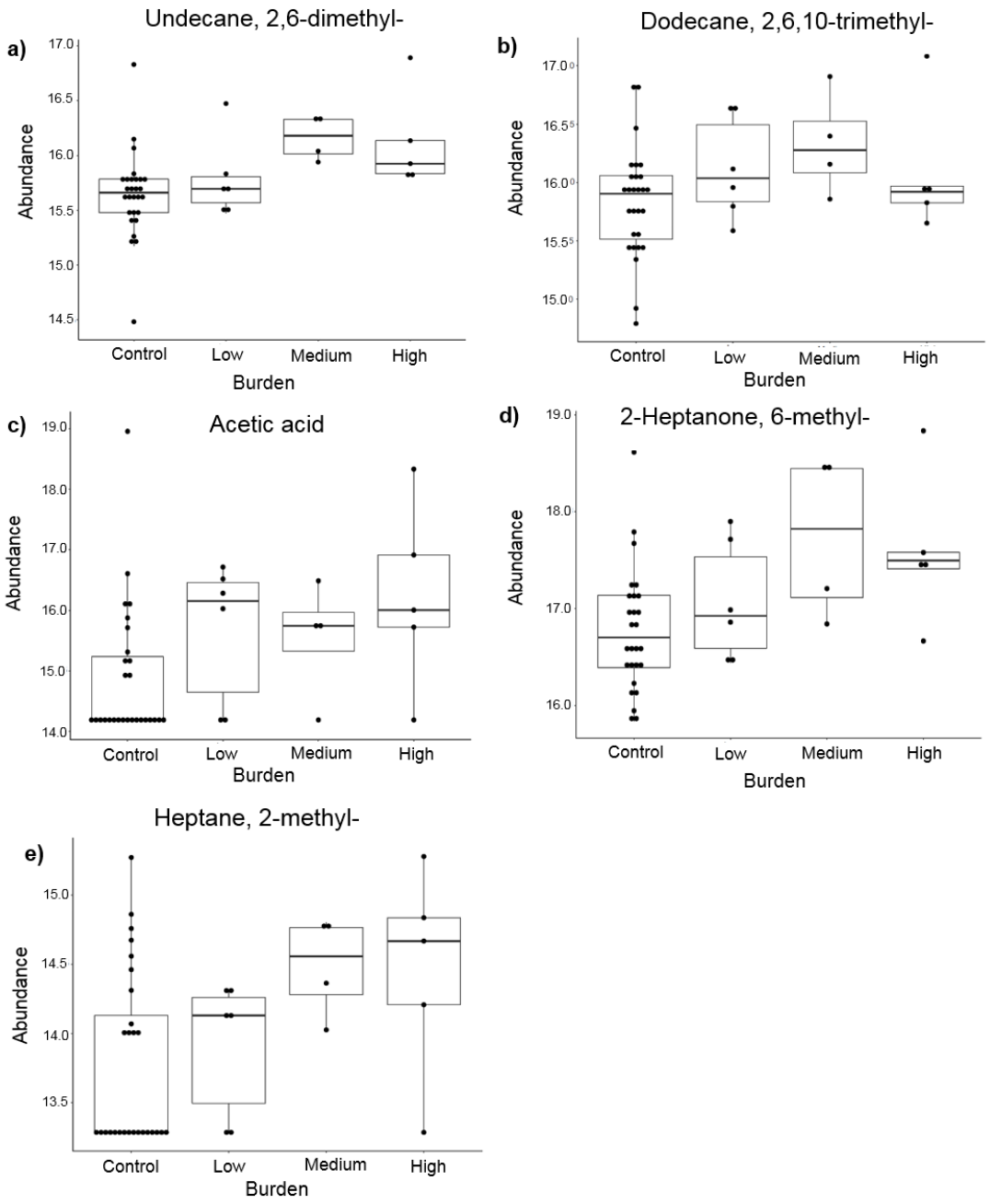


Figure 5.11 Box plots of VOC abundance of the colonic contents of horses infected with *A. perfoliata* and non-infected controls. The 3 VOCs with the highest statistical significance, identified by multivariate regression modelling for both AT vs Co (**a**-**c**) and MH vs CO (**a**, **d**, **e**) are shown. Box Plots were constructed to show VOC abundance change in individual groups (control, low, medium and high tapeworm) to show gradient change of compounds with level of tapeworm burden. Key: AT = all tapeworm samples, MH = tapeworm samples with 21+ worms, CO = control (tapeworm negative).

Samples included in PERMANOVA model	Variable	R^2	p-value
AT (n=20) and CO (n=28)	Tapeworm burden	0.07	0.03*
	Strongyle FEC	0.01	0.68
	GCMS batch	0.04	0.13
MH (n=9) and CO (n=28)	Tapeworm burden	0.08	0.02*
	Strongyle FEC	0.01	0.92
Low tapeworm (n=6) and CO (n=28)	Tapeworm burden	0.03	0.35
Medium tapeworm (n=4) and CO (n= 28)		0.06	0.07
High tapeworm (n=5) and CO (n=28)		0.07	0.04*
LSC FEC (n=10) and HSC FEC (n=24)	Strongyle FEC	0.02	0.53

Table 5.7 Results of PERMANOVA analysis of the VOC abundance of colonic contents of horses infected with *A. perfoliata* and non-infected controls. Key: AT = all tapeworm samples, MH = tapeworm samples with 21+ worms, CO = control (tapeworm negative), FEC = faecal egg count, LSC = low strongyles, HSC = high strongyles.

5.3.5 Integration of microbiome and metabolome data

From cross-validation of the model, a performance plot was constructed displaying the overall classification error rate and BER (Figure 5.12). Figure 5.13 shows component 1 and component 2 of a Pearson's correlation plot to show how well bacterial OTUs and VOCs can discriminate between AT and CO. Subtle clustering of AT and CO is shown (Figure 5.13a and b), with bacteria accounting for the greatest separator of the two omics (Figure 5.13c and d). The loadings for each component (demonstrating which variables are important in each group) are shown in Figure 5.14 a (bacteria) and b (VOCs). In Figure 5.15a is a circle plot which shows correlations between OTUs and VOCs. Clusters of points indicate strong correlations between variables. Based on the circle plot, a circos plot was constructed to visualise specific correlations between OTUs and VOCs as well as the groups they were more or less abundant in (Figure 5.15b). For a clearer representation of the circos plot a heatmap was also produced in Figure 5.15c to show positive and negative correlations (0.3 and above) between variables. Labels of OTUs included in the loadings for the Pearson's correlation plot (Figure 5.14a) and for the circos plot (Figure 5.15a) are in Appendix 5.7.

Correlations between specific VOCs and OTUs were also generated from the Pearson's correlation plot. A total of 1257 correlations (0.3 and above) were identified between OTUs and VOCs. To assess the use of VOCs as markers for OTUs in higher or lower abundance in

control or tapeworm samples identified in single omics, a table of VOCs and OTUs which were significant in the single omics analysis and had significant correlations in the integrated analysis was constructed (Table 5. 8). Of interest, OTU2331 and OTU147 (both Prevotellaceae) were in higher abundance in AT in single omics analysis and each positively correlated (0.36-0.6) with 19 VOCs which were also in significantly greater abundance in AT in single omics analysis. OTU2051 (Rikenellaceae) identified in higher abundance in the CO group was found to be negatively correlated (-0.6) with furan,-2-pentyl and 5hepten-2-one-6-methyl.

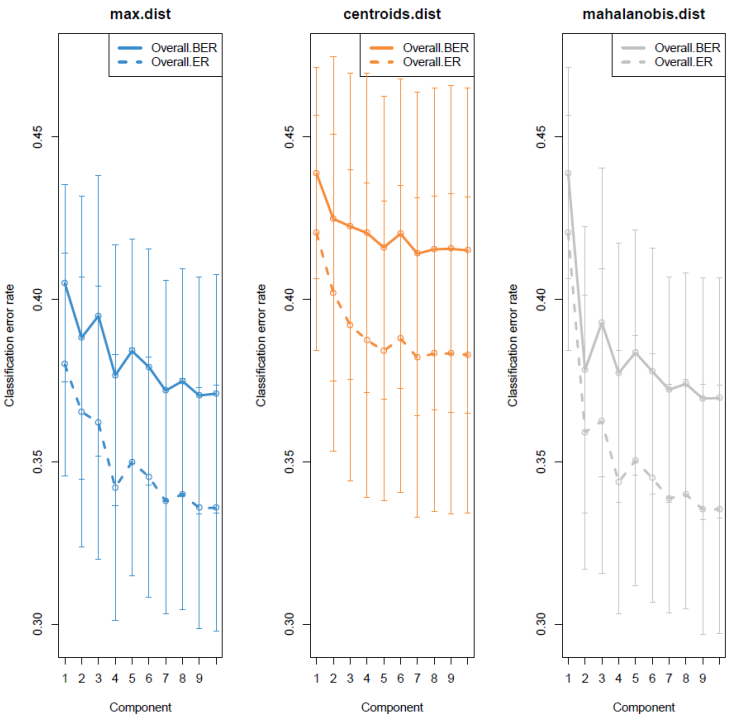


Figure 5.12 The classification error rate of a model used to combine bacteria and VOC data of the colonic contents of horses infected with *A. perfoliata* and non-infected controls. Key: ER = error rate, BER = balanced error rate.

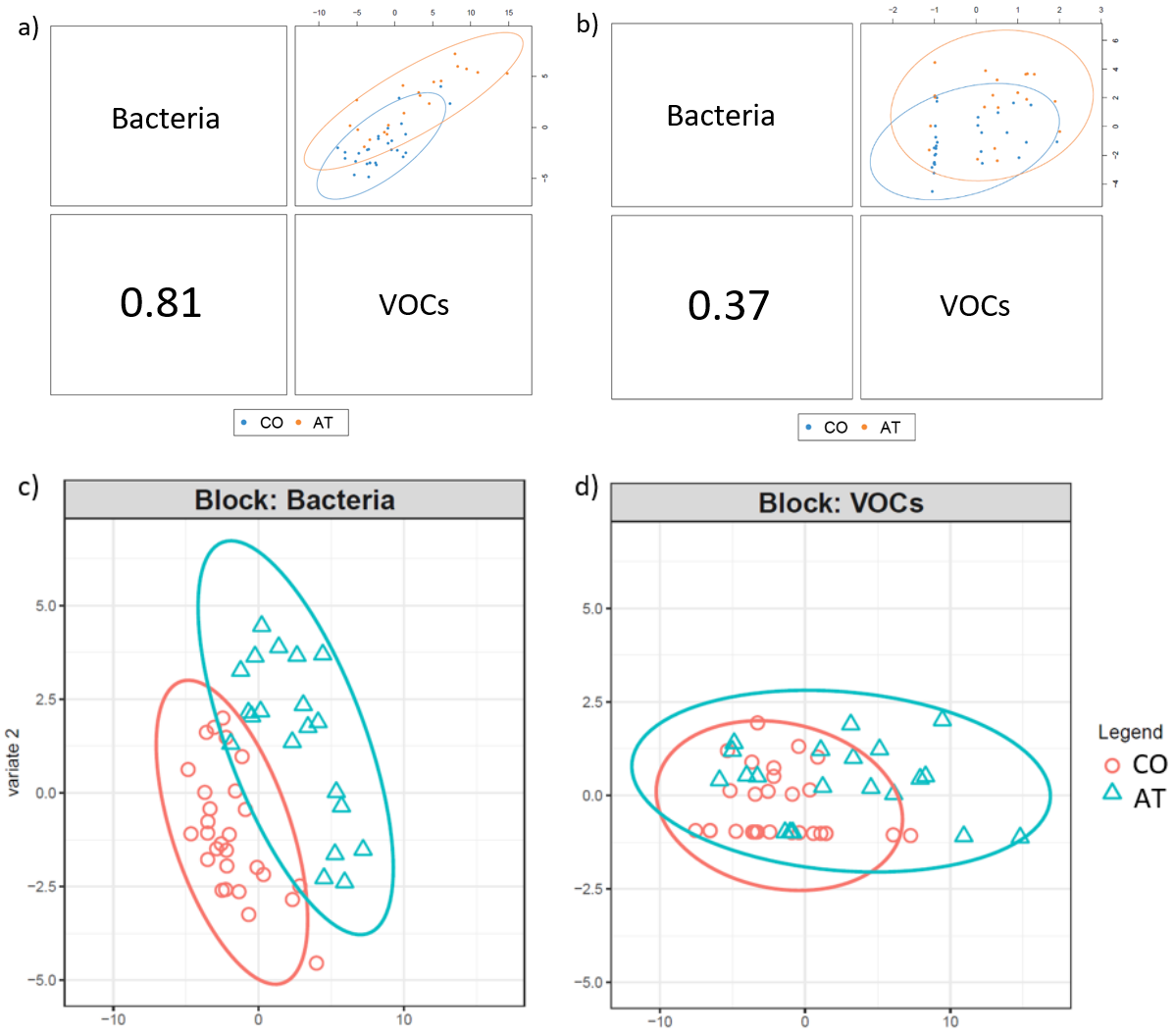


Figure 5.13 A Pearson's correlation plot of bacteria and VOC data of the colonic contents of horses infected with *A. perfoliata* and non-infected controls. Component 1 is shown in **a** and component 2 is shown in **b**. The ability of the model to separate CO and AT by bacteria alone is demonstrated by plot **c** and VOCs alone by plot **d**. Key: AT = all tapeworm samples, CO = control (tapeworm negative).

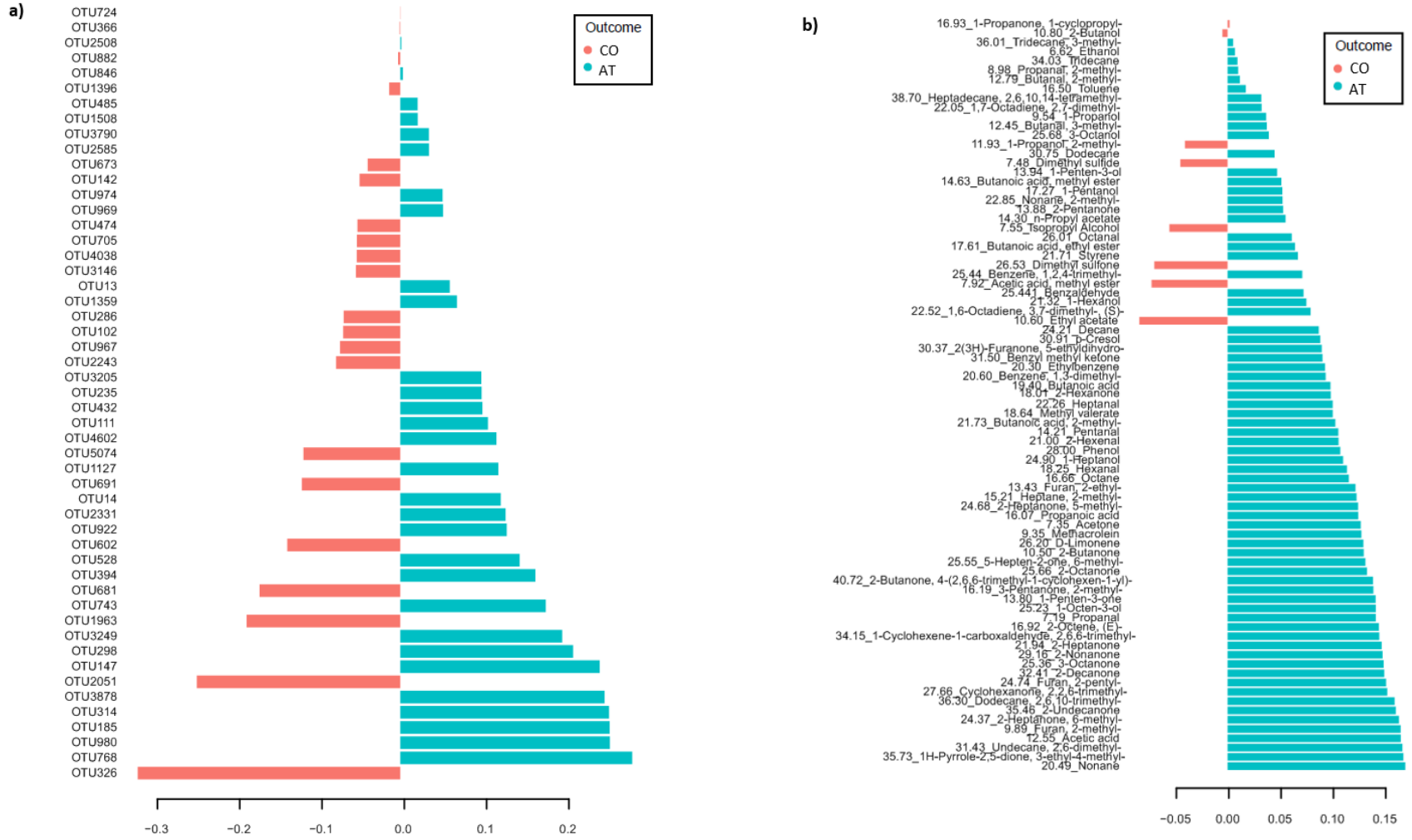
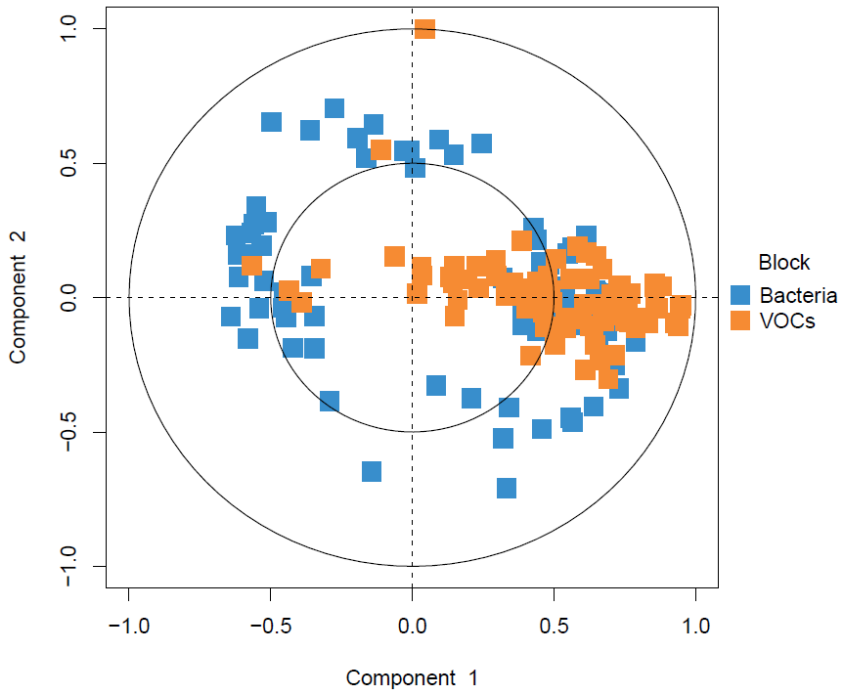
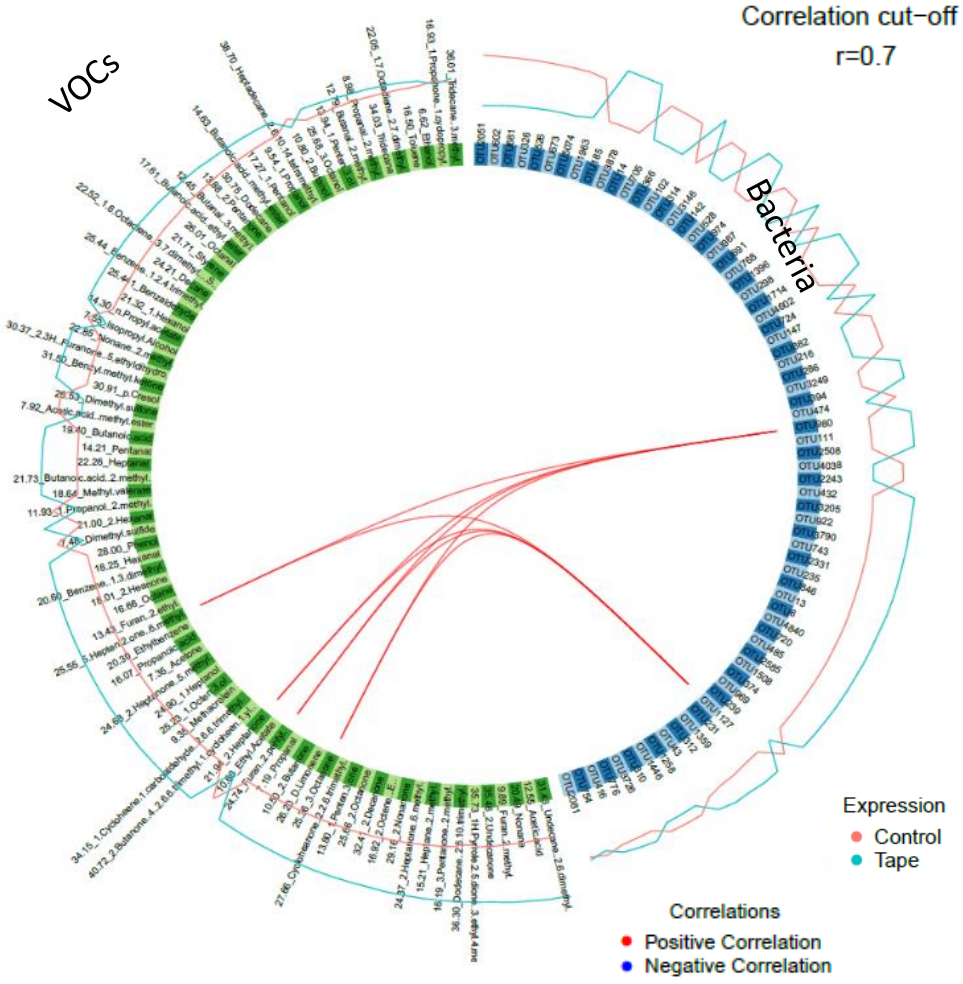


Figure 5.14 Loadings for component 1 of the Pearson's correlation plot (Figure 5.13a) of bacteria and VOC data of the colonic contents of horses infected with *A. perfoliata* and non-infected controls. OTUs of importance in component 1 are shown in **a** OTU labels can be found in Appendix 5.7. In **b**, VOCs important in component 1 are shown. Key: AT = all tapeworm samples, CO = control (tapeworm negative).

a)


Component 2
Component 1
1.0
0.5
0.0
−0.5
−1.0
Block
Bacteria
VOCs

b)


Correlation cut−off
r=0.7
VOCs
Bacteria
Expression
Control
Tape
Correlations
Positive Correlation
Negative Correlation

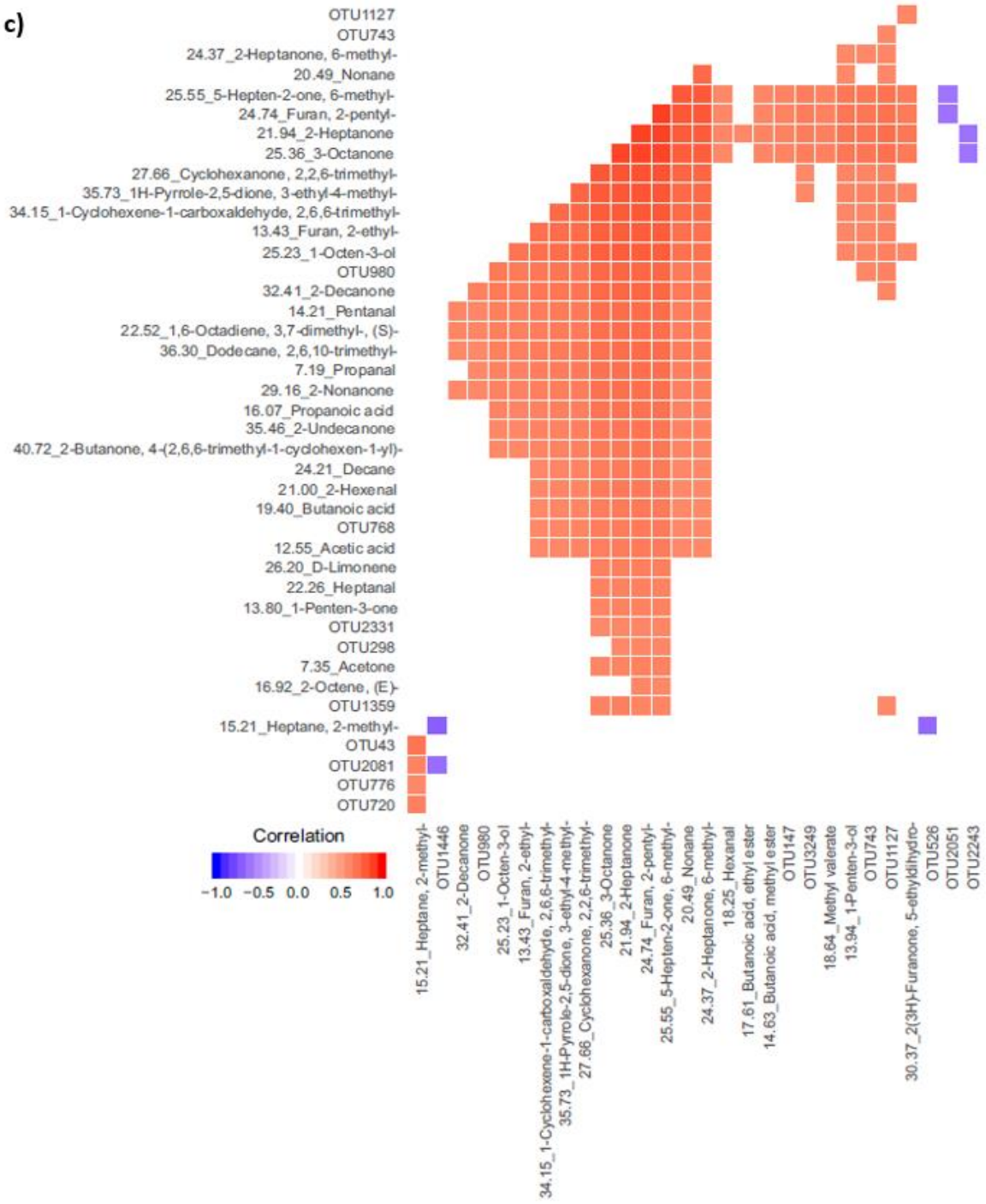


Figure 5.15 Correlation plots built from a model used to combine bacteria and VOC data of the colonic contents of horses infected with *A. perfoliata* and non-infected controls. A circle plot is shown in **a)** represents visualisation of positive and negative correlation between features. A circos plot to visualise specific OTU and VOC correlations generated by the circle plot is shown in **b)** OTU680 and OTU1127 belong to the family Lachnospiraceae. The correlation cut-off is 0.7. In **c)** a heatmap showing correlations of features from the circle plot with a lower cut off (0.3 and above). Labels of OTUs are in Appendix 5.7.

	P-value	OTU higher abundance in	VOCs significantly correlated	Correlation	VOC higher abundance in
OTU5074	<0.001	Control	Acetic acid	-0.42	AT
OTU3878	<0.001	Tape	2-nonanone	0.46	AT
			2-decanone	0.43	AT
			1-octen-3-ol	0.43	AT
OTU705	<0.001	Control	2-octanone	-0.46	AT
			1-octen-3-ol	-0.47	AT
OTU314	<0.001	Tape	1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	0.43	AT
			Cyclohexanone, 2,2,6-trimethyl-	0.41	AT
			Undecane, 2,6-dimethyl-	0.41	AT
			2-Heptanone, 6-methyl-	0.4	AT
			1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.4	AT
			Dodecane, 2,6,10-trimethyl-	0.37	AT
			Acetic acid	0.37	AT
OTU673	<0.001	Control	Heptane, 2-methyl-	-0.37	AT
OTU2331	<0.001	Tape	2-Heptanone, 6-methyl-	0.6	AT
			Decane	0.58	AT
			Propanal	0.56	AT
			1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.55	AT
			Furan, 2-pentyl-	0.55	AT
			Cyclohexanone, 2,2,6-trimethyl-	0.53	AT
			Undecane, 2,6-dimethyl-	0.53	AT
			1-octen-3-ol	0.51	AT
			1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	0.48	AT
			2-Heptanone	0.45	AT
			Pentanal	0.45	AT
			2-Decanone	0.43	AT
			Furan, 2-methyl-	0.43	AT
			3-Octanone	0.42	AT
			2-nonanone	0.4	AT
			2-Undecanone	0.4	AT
			Methacrolein	0.38	AT
			1-Heptanol	0.38	AT
			Nonane	0.37	AT
OTU2051	<0.001	Control	Acetic acid	-0.37	AT
			1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	-0.38	AT
			2-Heptanone	-0.39	AT
			1-octen-3-ol	-0.41	AT
			2-decanone	-0.45	AT
			2-octanone	-0.47	AT
OTU142	0.015	Control	2-nonanone	-0.37	AT
			Furan, 2-pentyl-	-0.38	AT

			2-Butanone, 4-(2,6,6-trimethyl-1-cyclohexen-1-yl)-	-0.4	AT
			2-decanone	-0.43	AT
			Propanal	-0.43	AT
			2-heptanone	-0.43	AT
			1-octen-3-ol	-0.44	AT
OTU147	0.015	Tape	2-heptanone	0.56	AT
			3-Octanone	0.53	AT
			1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	0.53	AT
			Furan, 2-pentyl-	0.53	AT
			2-decanone	0.51	AT
			2-nonanone	0.51	AT
			1-penten-3-one	0.49	AT
			nonane	0.48	AT
			2-Heptanone, 6-methyl-	0.47	AT
			2-Undecanone	0.47	AT
			1-octen-3-ol	0.45	AT
			Cyclohexanone, 2,2,6-trimethyl-	0.43	AT
			2-octanone	0.43	AT
			1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	0.42	AT
			Propanal	0.42	AT
			Furan, 2-methyl-	0.41	AT
			Dodecane, 2,6,10-trimethyl-	0.4	AT
			Acetic acid	0.39	AT
			D-limonene	0.36	AT
OTU239	0.020	Control	Decane	0.41	AT
			2-Heptanone, 6-methyl-	0.39	AT

Table 5.8 OTUs and VOCs identified in single omics analysis which were significantly correlated with each other when integrated. OTUs were identified as significantly different between CO and AT by DeSeq2 R package. VOCs were identified as significantly different between CO and AT by multivariate multiple regression. Correlations between OTUs and VOCs were recorded from 0.3 to 1 (-0.3 to -1) to indicate positive (negative) linear relationships. The correlations were calculated from a Pearson's correlation plot and the cor.test function to determine if correlations were statistically significant. OTU classifications: OTU5074 Treponema 2, OTU3878 Alloprevotella, OTU705 Bacteroidales S24-7 group, OTU314 [Eubacterium] oxidoreducens group, OTU673 Lachnospiraceae NK4A136 group, OTU2331 Prevotellaceae UCG-003, OTU2051 Rikenellaceae RC9 gut group, OTU142 Ruminococcaceae UCG-005, OTU147 Prevotellaceae UCG-001, OTU239 Prevotella 1.

5.4 Discussion

Comparisons of the colonic microbiome and metabolome of *A. perfoliata* infected and non-infected horses have not previously been investigated. For the first time differences in the gut microbiome attributable to parasitic helminths have been established. A reduction in symbiotic bacteria was observed in tapeworm samples which could have occurred as a result of parasitic infection, as previously observed by others (Clark *et al.*, 2018). Furthermore, integration of microbiome and metabolome data in relation to parasite status of the horse has not been previously published. In this study some statistically significant VOCs and OTUs identified in single omics analysis had moderate to strong correlations when data were integrated, indicating VOCs have the potential to act as markers for bacterial OTUs in equine colon contents. Overall, this pilot study warrants further investigation of the gut microbiome and metabolome in relation to parasites in horses and other animal species.

5.4.1 Parasitology results and category choice for analysis

The low category of tapeworm infection was defined based on previous research that 1-20 tapeworms have minimal impact on health (Pearson *et al.*, 1993). Therefore, $\geq$ 21 tapeworms are potentially clinically important. Studying such a clinically important group may give a better insight into the potential role of the gut microbiota in tapeworm-related colic. The thresholds set for the medium and high categories were arbitrary; they were chosen to demonstrate patterns with increased worm burden as there is a positive correlation between infection intensity and the risk of tapeworm associated colic (Proudman *et al.*, 1998). Furthermore, in a population of horses studied by Proudman & Trees (1996b), few horses had >100 tapeworms, therefore it was considered sensible that a threshold of $\geq$50 would be suitable for high burdens in this work.

The FECs for the horses in this study were generally quite low, with most samples (75%) falling into a low category, as defined the AAEP guidelines and work by Nielsen and colleagues (Kaplan & Nielsen, 2010; Nielsen *et al.*, 2010b, 2013). There were no samples in which the FECs were high and 25% of samples were defined as medium burdens. Anthelmintic treatment history and the age of the animals in this study were unknown – both factors have previously been identified to affect egg shedding (Lester *et al.*, 2018). Furthermore, spring and summer have been reported as times of year most likely to yield high FECs (Wood *et al.*, 2013). Horses in this study were

sampled during the autumn which could further explain the lower counts observed. The concept of using a FEC to determine whether a horse requires treatment for stongyles has been implemented since the early 1990s (Duncan & Love, 1991). However, the FEC test is not an accurate indicator of strongyle burden, particularly during the autumn as up to 90% of strongyles are encysted during this time and are not shedding eggs (Murphy & Love, 1997). For future studies investigating tapeworm infection on the gut microbiota and metabolome, careful consideration of how to control for strongyle burdens must be made. In live animals, treatment with an effective anthelmintic against encysted strongyles (e.g. moxidectin) and withheld grazing (to prevent re-infection) may be a strategy to control for the presence of strongyle species when investigating the effect of tapeworm alone (Lyons *et al.*, 2017; Lester *et al.*, 2018).

5.4.2 Freeze-dried compared to non-freeze-dried colon contents

The water content of equine colon contents has been reported to be up to 90%, whereas faecal dry matter can range between 20 and 30% for pastured and stabled horses, respectively (Sneddon & Argenzio, 1998; Williams *et al.*, 2015). A study by Aggio *et al.,* (2016) found that freeze-dried human urine samples yielded greater numbers of VOCs than non-freeze-dried urine. The authors proposed that freeze-drying reduced the liquid phase and increased the gas phase making a greater number of VOCs available in the headspace for absorption by the SPME fibre (Aggio *et al.*, 2016). In the current work, freeze-dried equine colon contents also resulted in an increase in VOCs. In a comparison of weights of wet faeces in **Chapter 2**, significantly fewer VOCs were observed in 100 mg than 1000 mg. Here, despite the fact 100 mg of freeze-dried material was compared to 1000 mg of non-freeze-dried material significantly more VOCs was observed in the freeze-dried material. A reduction in the water peak of freeze-dried samples may have allowed the relative proportions of other compound peaks to increase and therefore be identified more easily (Figure 5.7). However, in human faeces VOC profiles of diluted and non-diluted samples did not differ significantly from each other (Berkhout *et al.*, 2016). Additionally, others have demonstrated that freeze-dried human faeces had a reduced number of compounds compared to wet faeces (Phua *et al.*, 2013). The different outcomes observed after freeze-drying equine colon contents and human urine compared to human faeces may be because of the unique chemical composition of these biological matrices or the metabolomics platform used. A published method to extract VOCs from equine colon contents is not currently available - further studies are necessary to accurately compare wet and dry methods of extracting VOCs from

equine colon contents to determine the optimal weight for use in GCMS studies. Furthermore, despite yielding more VOCs by freeze-drying, the more volatile compounds may have been lost by the freeze-drying process and therefore may not be representative of the native sample. However, in this work the main microbial compounds of interest (SCFAs) were increased in the freeze-dried colon contents therefore the use of freeze-dried material was justified.

5.4.3 A comparison of the intestinal bacterial microbiome and metabolome of tapeworm infected and non-infected horses

Firmicutes (51.6%) and Bacteroidetes (36.1%) were the dominant phyla in this work, which agrees with other populations of horses studied (Steelman *et al.*, 2012; Weese *et al.*, 2014; Clark *et al.*, 2018). The bacterial diversity between tapeworm and CO samples were similar, a finding consistent with many other studies comparing the gut microbiome of parasite infected and non-infected animals (Li *et al.*, 2011b, 2016; Šlapeta *et al.*, 2015; Duarte *et al.*, 2016; Clark *et al.*, 2018; Peachey *et al.*, 2018). Significant differences in microbiota diversity in parasite positive and negative animals have been more widely reported in rodent and rabbit studies in which the level of infection and diet can be closely controlled and larger sample sizes can be used (Houlden *et al.*, 2015; Arrazuria *et al.*, 2016; Cattadori *et al.*, 2016; Jenkins *et al.*, 2018). One murine study found the effect of natural tapeworm on the gut microbiota was evident upstream, as well as downstream in the intestinal tract (Kreisinger *et al.*, 2015). Upstream impact may indicate a systemic effect of parasites on the gut microbiota interacting with the host and vice versa. The region sampled in this work was the pelvic flexure, downstream from the location of tapeworms in the gut (ileo-caecal junction and caecal wall). Sampling numerous locations of intestinal contents in future studies could determine the wider impact of tapeworm infection on the gut microbiota.

Increased members of the phylum Proteobacteria were observed in horses with high or a susceptibility to a high strongyle FEC by Peachey *et al*., (2018) and Clark *et al*., (2018), but this was not observed in tapeworm infected horses or in those in the HSC group in the current work despite a larger sample size. Possibly the levels of infection were not high enough to observe differences in members of Proteobacteria between groups or mixed infections and a lack of control for diet in the current study may have been all been confounding factors. Here, a down-regulation of some symbiotic bacteria, including an unidentified genus belonging to the family Ruminococcaceae UCG-004 was observed in the tapeworm groups when compared to the

controls. In other species, fewer Ruminococcaceae were seen in the rabbit and the pig (genus level, *Ruminococcus*) infected with intestinal parasites (Wu *et al.*, 2012; Cattadori *et al.*, 2016). A decrease in *Ruminococcus* was previously identified as a consequence of strongyle infection in ponies identified as being susceptible (Clark *et al.,* 2018). The authors of the latter study suggested a reduction in butanoic acid (because of a reduction in *Ruminococcus*) could influence the level of inflammation caused by infection. However, Clark and colleagues did not measure butanoic acid and in the current work a correlation between butanoic acid and symbiotic bacteria was not observed. The role of symbiotic bacteria during parasite infection requires further investigation to determine if interactions occur as a result of infection or as a pre-cursor and what consequences this may have for the host.

In the MH group the genus *Selenomonas 3* was more abundant compared to CO. Previously *Selenomonas ruminatum* was found to be more abundant in goats infected with *Haemonchus contortus* which shares the mucosa-feeding trait with tapeworm (Li *et al.*, 2016). Some *Selenomonas* species have been associated with inflammation in patients with periodontal disease (Nagpal *et al.*, 2016), therefore *Selenomonas* may be associated with inflammation caused by mucosa-feeding parasites in goats as observed by Li and colleagues and horses as observed here. Equally, species of *Selenomonas* are involved in the fermentation pathway of starch and sugars in the hindgut so an increased presence may be attributable to the diet (Shirazi-Beechey, 2008) or to an adaption of the host or gut microbiota to compete for nutrients with the parasite. In the current study, just two out of nine horses in the MH group had high levels of *Selenomonas 3*. In general, as tapeworm burden increases, severity of lesions, inflammation of the mucosa and the risk of colic (intussceptions, perforation and impactions) increases (Pearson *et al.*, 1993; Proudman *et al.*, 1998; Kjaer *et al.*, 2007; Hreinsdóttir *et al.*, 2019). However, fewer tapeworms are required to evoke severe lesions at the ileo-caecal junction than the caecal wall (Williamson *et al.*, 1997). Therefore, numbers of tapeworm may not have been fully representative of mucosa inflammation in this work and the specific site and extent of inflammation was not recorded. It cannot be concluded that the two horses with more abundant *Selenomonas 3* had particularly inflamed mucosa or that the effect was unrelated to tapeworm infection. Furthermore, species level identification would have provided further clues about the function of the *Selenomonas* organisms, but this was beyond the scope of this study. Further investigation, recording sites and levels of inflammation by taking biopsies to study mucosa histology and the microbiota at inflamed and non-inflamed sites would be more indicative of

species related to inflammation. The role of the gut microbiota and tapeworm in the development of or protection against equine colic remains unclear. In terms of protection against inflammation evidence in mice has shown those with tapeworm or nematode infections had lower chemically-induced (dextran sulfate sodium or dinitrobenzene sulfonicacid, respectively) colitis than tapeworm or nematode negative mice (Reardon *et al.*, 2001; Khan *et al.*, 2002).

A PERMANOVA analysis revealed tapeworm burden explained 7-8% of the variation in VOC profile, but evidently, this was not enough to demonstrate clear clustering for tapeworm in PCA analysis. The only other variable measured in this study was strongyle FEC, which was not associated with clustering in a PCA nor did it explain variation in the VOC profile by PERMANOVA. In general, strongyle FEC had less of an impact on VOCs and the gut microbiota than tapeworm. This could have reflected the relatively low FECs recorded. Other factors, discussed in further detail in **Chapter 1** which have been previously described to influence the equine hindgut microbiome and metabolome including diet, gastrointestinal disease and horse type (Yamano *et al.*, 2008; Daly *et al.*, 2012; Dougal *et al.*, 2014) (not recorded here) may have been responsible for the rest of the variation observed in the data. For future investigations, to maximise the ability to determine variation influenced by parasites alone on the microbiota, extrinsic factors (diet, gastrointestinal disease, horse type etc.) should be controlled for. However, this information may be difficult to obtain for horses submitted to an abattoir.

The VOCs of interest identified in this work were mostly in greater abundance in the tapeworm groups. Among these VOCs were undecane, 2,6-dimethyl-; octen-3-ol; 2-octanone and furan compounds, which are of potential fungal origin (Effmert *et al.*, 2012b). Many of the 'fungal' VOCs identified here correlated with fungal OTUs in **Chapter 3** and are discussed there in further detail. However, there are no data published on the interaction of gut fungal populations with tapeworm parasites specifically. Previously, ponies susceptible to strongyle infection had higher concentrations of anaerobic fungi at 3 time points out of 5 over a 4 month trial (Clark *et al.*, 2018). However, Clark and colleagues did not investigate the relationship of the abundance of anaerobic fungi specifically with parasite burden. Furthermore, many of the VOCs related to the tapeworm groups in this chapter were identified as highly correlated with facultative-anaerobic fungi in **Chapter 3**.

The horses used in this study were sampled in the autumn which aligns with the work in **Chapter 3** where a marked increase in facultative-anaerobic fungi, and correlating potential fungal VOCs,

was observed in the autumn. It may be reasonable to consider that the tapeworm infected horses sampled in this work were more likely to have been grazing on pasture prior to death than the non-infected controls. It has been established previously that horses grazing at pasture are more likely to become infected with tapeworm than non-grazing horses (Trotz-Williams *et al.*, 2008). While grazing the animals may have picked up both oribatid mites and fungi, whereas the controls may have had fewer fungal VOCs because they were simply not grazing on pasture. However, whether the horses had access to grazing during the autumn before death was not recorded in this study and further work would be needed to confirm this hypothesis. In terms of the impact of this phenomenon on health, many of these fungal VOCs were correlated with two OTUs (also more abundant in the tapeworm group) belonging to the family Prevotellaceae. Members of the Prevotellaceae family have been associated with driving chronic intestinal inflammation in mice with inflammasome-mediated dysbiosis (Elinav *et al.*, 2011; Scher *et al.*, 2013). It may be reasonable to speculate that the higher abundant OTUs were associated with parasite-driven inflammation. Interaction of the gut microbiota and parasites in terms of bacteria and development with inflammation in veterinary species has been discussed (Peachey *et al.*, 2017), but whether interactions exist between gut fungi and parasites and the impact on the host is unknown. On the other hand, in healthy human subjects consuming a high-carbohydrate diet, a correlation was observed between the fungal genus *Candida* and members of Prevotellaceae (genus *Prevotella)* (Hoffmann *et al.*, 2013). The authors speculated that *Candida* were able to break-down complex carbohydrates into simple sugars to be fermented by *Prevotella*, to produce acetic acid. In the current work acetic acid was also more abundant in the tapeworm group, further supporting this particular hypothesis. In this study the OTUs of interest could not specifically be identified as Prevotella, but as members of the family Prevotellaceae, in which there as three other possible genera the OTUs may have belonged to. Fungal populations were not characterised in this study so it can only be speculated that these VOCs were related to fungal metabolism. An environmental fungus (*Duddingtonia flagrans*) has demonstrated properties for the control of nematode parasites (Larsen *et al.*, 1995; Braga *et al.*, 2009). This is yet be investigated in cestode species. However, the intermediate host of tapeworm (oribatid mite) is a potential vector for some fungal taxa, including members of Ascomycota, Basidiomycota and Zygomycota (Renker *et al.*, 2005). Therefore, ingestion of the oribatid mite may be a source of both parasitic infection and fungi. Further studies are required to sample

abattoir material at different times of the year for a clearer indication of whether fungal VOCs were related to the time of year or parasite status.

In the single omics analysis 69 OTUs were found to be significantly different between AT and CO. Of these, 10 were correlated with VOCs (listed in Table 5.8). A lack of correlation for the other 59 OTUs with VOCs could imply the organisms were not active, as DNA-based studies are not representative of active species (Schirmer *et al.*, 2018). Furthermore, the use of one metabolomics platform and extraction technique will not encompass all metabolites within a biological matrix (de Raad *et al.*, 2016a) so there is a possibility that metabolites directly correlating with the most abundant organisms were not detected here. The use of multiple metabolomics platforms for gut microbial related studies in the horse would provide more information for future studies.

VOCs were not significantly different in abundance between tapeworm infected and non-infected controls after p-values were adjusted for multiple comparisons. Despite not reaching statistical significance in the single omics analysis many of these VOCs were responsible for the variation shown in the Pearson's correlation plot (supervised analysis) and were correlated with several OTUs (Table 5.8, Figures 4.14 and 5.15). This demonstrates that, although statistical significance was not reached after correction for multiple comparisons, these VOCs were still important in discriminating between tapeworm infected and non-infected controls when combined with other analysis and the study may have simply been under-powered when evaluated as a single omics. This also demonstrates the advantages of using supervised analysis in datasets such as this which may be particularly noisy because of confounding factors (Gromski *et al.*, 2015). However, when supervised analysis is employed it is important to check that the data are not being over fitted. In the present study the BER was used to evaluate the supervised model for overfitting. Here, a BER of 35-45% for the mixOmics model was rather high for all distances measured. Bijlsma *et al.,* (2016) proposed an accepted error rate of lower than 30% was able to demonstrate true differences, whereas an error rate of 30-40% was indicative of possible differences between groups. Therefore, the error rate in this work is an indication that the group separations of AT and CO in Figure 5.13 may indicate possible differences between groups but should be interpreted with caution. The high error rate may be because of the small sample size (more features than samples were included in the model) or because the differences between the two groups were not enough to accurately classify the samples. In the horse, this is

among the first work to correlate findings of a 16S rRNA study with the gut metabolome in a statistical model (also performed in **Chapter 3**). The integration of omics to address the same hypothesis was able to provide stronger evidence for the conclusions made here. A larger sample size is required to validate whether the combination of bacterial OTUs and VOCs can accurately distinguish between horses with and without tapeworm infection.

5.4.4 A comparison of the intestinal bacterial microbiome and metabolome of horses with high and low strongyle FECs

A comparison between horses with high and low strongyle FECs was performed to see if the results could replicate the findings of Peachey *et al.,* (2018). Peachey and colleagues controlled for diet (all horses were grazing at pasture) and controlled for tapeworm parasites (all horses received a dose of Praziquantel before the trial). Two consistencies between the previous work and the current study were found. Firstly, there was no significant effect of strongyle FEC on the overall gut or faecal microbiota diversity. Secondly, members of Methanomicrobia (archaea) were found to be more abundant in the low strongyle FEC group in both studies. This was observed at class level by Peachey *et al*., (2018) in contrast to OTU level in the current work: two OTUs belonging to Methanomicrobia (OTU264 Methanocorpusculum and OTU388 Methanobrevibacter) were more abundant in the low strongyle group. Class level differences may not have been observed in the current work because of the confounding factors of mixed parasite burden and lack of control for diet. These findings have relevance for future trials studying the microbiota of horses (strongyle FEC appears to have little association with overall gut or faecal diversity). The up-regulation of members of Methanomicrobia may be a signature in the gut and faeces of horses with low strongyle FECs, although this finding was not repeated in a recent study in a group of yearlings (Peachey *et al.*, 2019). The importance of an increased abundance of Methanomicrobia members in context with low strongyle burdens requires further investigation. Peachey *et al.,* (2018) proposed that methanogens (such as Methanomicrobia) were able to elicit TH17 immune responses to clear parasite infection (Reynolds *et al.*, 2014b; Bernatchez *et al.*, 2017). Equally, as with the current work it could not be ruled out that grazing behaviour could have influenced a coincidental link between low strongyles FEC and higher copies of Methanomicrobia genes.

Unfortunately, in this work it was not possible to compare differences in the microbiota between horses with high levels of mixed parasite burdens with those with low mixed burdens because of

small sample sizes. Investigating the impact of general high burdens on the equine microbiota and overall health would be an interesting observation and add to further knowledge of parasite-microbiota interactions in the horse.

5.5.5 Overall discussion and conclusions

The overall aim of this chapter is to compare the colonic microbiome and metabolome of tapeworm infected and non-infected horses. The study design allowed horses to be diagnosed for tapeworm by gold standard (counting of worms in the caecum) and for colon contents to be sampled rather than faeces. To meet these criteria sampling had to be performed post-mortem. The most straightforward method to sample horses in this way is via abattoir material. However, this meant that information on previous diet, gastrointestinal disease and horse type could not be recorded. The lack of control for extrinsic factors is the main limitation of this study. The sampling of abattoir material does not allow repeated measurements (e.g. before and after exposure to parasites), hence this study was unable to determine whether differences observed between tapeworm infected and non-infected controls were a consequence or a precursor of parasite infection. The advantages and limitations of a marker-based method (16S rRNA) was discussed in **Chapter 2**, the use of this method was deemed appropriate for a pilot study.

Despite these limitations, differences between tapeworm infected and non-infected horses were observed and data collected for the microbiome and metabolome were successfully combined to identify correlations. Understanding the mechanisms responsible for microbiota, parasite and host interactions were beyond the scope of this study. Whether *A. perfoliata* infection and the gut microbiota interact to produce positive (stimulation of the immune system) or negative (a role in the development of colic) impacts on the host remains unknown. Others have proposed the idea that the use of probiotics to manipulate the gut microbiota may help control gastrointestinal parasites as an alternative to drugs. However, before further exploration is performed, the wider impacts of such therapies on the host should be considered. Research of the interactions between intestinal parasites and the microbiota is still very much in the early stages in the horse. The current work was the first pilot study to imply subtle differences in the microbiome and metabolome of horses infected with *A. perfoliata* and those negative for the parasite. This has demonstrated the study of the microbiome and metabolome in relation to *A. perfoliata* is a valid area of research and deserves further attention.

Chapter 6 An investigation of the use of VOCs from the headspace of rectal contents as biomarkers for *Anoplocephala perfoliata* (tapeworm infection) in horses.

6.1 Introduction

A. perfoliata (tapeworm) is a highly pathogenic and common equine gastrointestinal parasite (Proudman & Holdstock, 2000; Tomczuk *et al.*, 2014). As discussed in **Chapter 1** and **Chapter 5**, lessons learnt from growing anthelmintic resistance in strongyle populations have led to concerns regarding the use of the two drug classes currently available to treat tapeworm (Nielsen, 2016). Parasite testing with targeted treatment strategies has been proposed to slow down anthelmintic resistance in equine intestinal parasites, including for tapeworm (Rendle *et al.*, 2019).

A serological ELISA for tapeworm was developed in the 1990s, (Hoglund *et al.*, 1995; Proudman & Trees, 1996a). However, it requires a veterinary surgeon to take blood and the costs of this in addition to costs of laboratory analysis may be a barrier for some horse owners who may choose to administer anti-cestode treatments regardless of whether the horse has a tapeworm burden or not. An additional, more recently available diagnostic test for equine tapeworm (EquiSal®) is a saliva-based an antibody titre-based ELISA. This has been on the market since 2014 and has been shown to reduce the use of anthelmintics in a population of horses (Lightbody *et al.*, 2018). The use of a saliva sample is advantageous as it is non-invasive, although it does require horses to be fasted for 30 minutes prior to the collection of a sample, which may not always be practical (EquiSal user manual, Austin Davis Biologics). However, the cost of an EquiSal® test can exceed the price of a tapeworm treatment, which may not motivate horse owners to use it. The reported sensitivity (83%) and specificity (85%) values of the saliva ELISA have shown encouraging results, but they are not a perfect test (Lightbody *et al.*, 2016). Moreover, the use of antibody titre-based assays are not suitable for the monitoring of drug resistance because of slow antibody clearance in the absence of worms, which is not indicative of real-time infection (Abbott *et al.*, 2008). For these reasons further investigation into improved diagnostics that are easy to use, cost-effective and accurate is warranted.

In **Chapter 5** some differences in the colonic microbiome and VOC metabolome were observed between tapeworm infected and non-infected horses. Together with **Chapter 3**, strong correlations between microorganisms and VOCs suggest that VOCs have the potential to act as a proxy for microorganisms. However, the question remains as to whether VOC patterns observed in the distal regions of hindgut are accurately reflected in the faeces. To be able to monitor equine gut health in terms of the microbiota and parasites, a VOC biomarker that is detectable in the faeces would be beneficial. The first aim of this chapter is to compare the VOC profiles of colonic and rectal contents using a subset of horses from **Chapter 5**. Previously, non-invasive VOC markers have been identified for Giardia and Malaria parasites in humans as well as for schistomiasis in mouse models (Li *et al.*, 2011a; Bond *et al.*, 2015; De Moraes *et al.*, 2018). The second aim is to investigate a similar concept in horses by comparing the rectal contents from tapeworm-infected and non-infected horses to identify potential VOC biomarkers.

6.2 Methods

6.2.1 Horses

Rectal contents (~25 ml) were collected from horses slaughtered at an abattoir for non-experimental purposes in November 2015 (Batch 1, n=46) and November 2017 (Batch 2, n=33). The methods of collection, thresholds for defining *A. perfoliata* burdens and characteristics of the samples in Batch 1 were described in **Chapter 5**. For Batch 2, samples of rectal contents were collected and stored in the same way as batch 1. FECs were performed on Batch 2 samples using the same method as Batch 1 (**Chapter 5**).

6.2.2 Sample preparation and VOC profiling

Samples were prepared and GCMS analysis was performed as optimised in **Chapter 2**. Samples in Batch 1 were analysed on the GCMS in 2016 and Batch 2 samples were analysed in 2018. Three technical replicates of each sample were run. One of the aims of this chapter was to determine whether the VOC profile of rectal contents were representative of colon contents (collected at the pelvic flexure). To ensure that an accurate weight of colonic and rectal material could be directly compared, 100 mg of freeze-dried (FD) colon and rectal contents from 6 horses, 3 controls (CO) and 3 with that were tapeworm positive ($\geq$1 tapeworms) (AT) were analysed on the GCMS. The comparison between colonic and rectal contents was made on a sub-sample of 6

horses (T7, T15, T9, C6, C19 and C13), because of limited GCMS availability and small sample volumes left over.

6.2.3 Data processing and statistical analysis

Data processing and identification of VOCs was performed as described in **Chapter 2.** To evaluate differences in the VOC profile between the two gut regions, comparisons were made for both FD material and non-freeze-dried (NFD) material. Visualisation of comparisons of the VOC profile of rectal and colonic contents were made using PCA, chromatogram overlays and a stacked plot of chemical classes. For statistical differences in VOC abundance between groups, a paired t-test was performed. Comparisons were not made between FD and NFD as colonic samples were compared in **Chapter 5** and FD and NFD were analysed on separate GCMS runs, which may have introduced bias.

Samples were collected in two batches and clustering for GCMS batch was evident (Figure 6.1). Subsequently, for comparisons between controls and tapeworm infected horses, a batch correction tool using the function PerformBatchCorrection() in Metaboanalyst was used (Johnson & Li, 2007). Statistical comparisons were made between CO (n=52) and AT (n=27) as well as CO (n=52) and MH (≥21 tapeworms, n=16), for the same reasons described in **Chapter 5**.

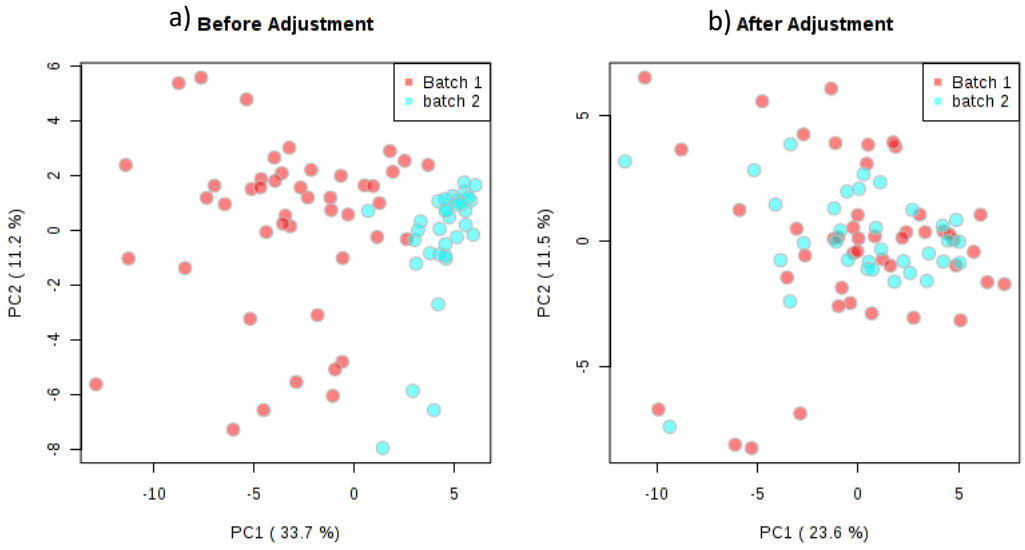


Figure 6.1 The VOC profile of rectal contents collected post-mortem from horses that were positive or negative for tapeworm infection. Samples were analysed on the GCMS in two batches the figure shows the VOC profiles pre- (a) and post-correction (b) for GCMS batch effect.

Initially an unsupervised (PCA) and supervised analyses (sPLS-DA) were performed in R to visually assess clustering for tapeworm and control groups and between the 2 groups. To evaluate significant differences in VOC abundance between groups, t-tests were performed. P-values were corrected for multiple comparisons using FDR. Selection of features for a ROC curve model was based on important features arising from the sPLS-DA model and using least absolute shrinkage and selection operator (LASSO). The ROC curve model was built using the algorithm logistic regression and was based on 80% of samples. A computer randomised set of samples representing 20% of the control group and 20% of the tapeworm groups was held back to test the model. The sensitivity and specificity values were calculated based on an average of the predicted class probabilities of each sample made across 100 cross-validations. P-values for ROC curves were calculated based on 1000 permutations. The permutation method was detailed in **Chapter 2**. T-tests and ROC curves were performed using the Metaboanalyst online software (Xia *et al.*, 2012).

As in **Chapter 5,** PCA plots were used to compare VOC profiles between low (0-199), medium (200-499) and high ($\geq$500) strongyle FECs. The criteria outlined by Peachey *et al.,* (2018) were also employed to determine if differences in FEC could be observed between LSC (0-10 e.p.g) and HSC ($\geq$200) samples.

6.3 Results

6.3.1 Parasitology

Numbers of tapeworms and FECs for horses in Batch 1 are listed in **Chapter 5.** Horses T2, T3, T8, T10 and T11 were not included in the current chapter because the level of tapeworm burden was unrecorded and may bias results when correcting for GCMS batch effect as these samples could not be matched to Batch 2. Tapeworm burdens and FEC results for horses sampled in batch 2 are recorded in Table 6.2.

Horse ID (Control)	FEC (e.p.g)	Horse ID (tapeworm positive)	*A. perfoliata* status	FEC (e.p.g)
C31	1	T22	High	393
C32	504	T23	High	165
C33	20	T24	Medium	5
C34	75	T25	Medium	123
C35	13	T26	Low	132
C36	144	T27	Low	0
C37	20	T28	Low	6
C38	110	T29	High	165
C39	75	T30	Low	144
C40	19	T31	Low	16
C41	47	T32	Medium	3
C42	96			
C43	381			
C44	1			
C45	0			
C46	95			
C47	63			
C48	16			
C49	14			
C50	96			
C51	46			
C54	8			

Table 6.1 Faecal egg count (FEC) results and *A. perfoliata* burdens of the horses included in Batch 2. Tapeworm burdens were recorded as low (1-20), medium (21-49) and high (≥50).

6.3.2 Comparison of the VOC profile of colon and rectal contents

PCA plots of rectal and colon samples are shown in Figure 6.2. In Figure 6.2a the samples were FD and in Figure 6.2b the samples were NFD. Regardless of whether the samples were FD or NFD, separation between colonic and rectal samples was evident, with two components describing 57% and 58.6% of the data for FD and NFD, respectively. A total of 83 VOCs were detected in NFD rectal and colonic contents; of these 16 (19.3%) were significantly different in abundance between colonic and rectal contents (Table 6.2). In the FD samples, 95 VOCs were detected and 14 (14.7%) were significantly different in abundance between colonic and rectal contents (Table 6.2). In NFD samples numbers of VOCs were significantly higher in the rectal contents than the colon (p=0.002, Table 6.3). Numbers of VOCs were not significantly different between colon and

rectal contents of FD samples (p=0.1, Table 6.3). Chromatogram overlays of FD and NFD colon and rectal samples from one horse (T7) are shown in Figure 6.3. A stacked plot of VOCs grouped for chemical class comparing colon and rectal samples when FD and NFD is shown in Figure 6.4. Chemical classes of VOCs were relatively similar, with some subtle differences across colon and rectal samples. In both FD and NFD, there were more alcohol compounds observed in the colon than the rectal contents. In contrast, there were more acids and esters in the rectal than the colon contents.

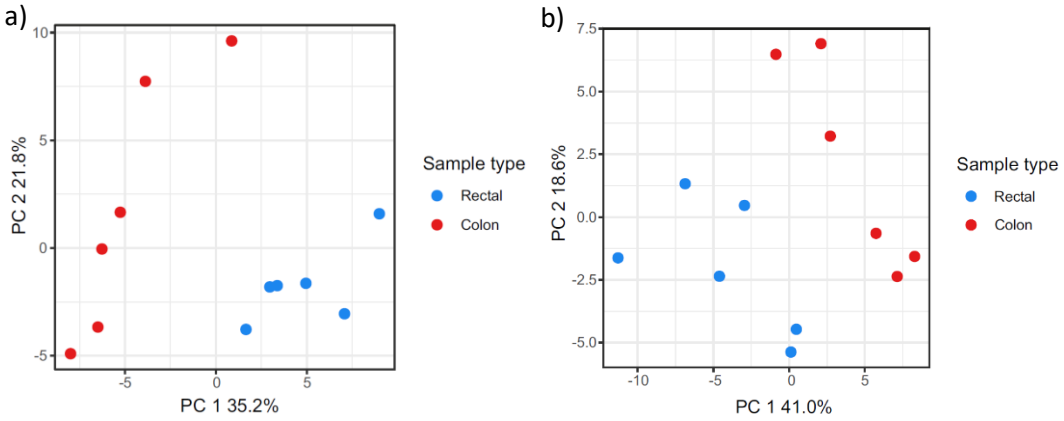


Figure 6.2 PCA plots of the VOC profiles of the colon and rectal contents of 6 horses. In **a)** the samples (each 100 mg) were freeze-dried prior to analysis and in **b)** the samples (each 1000 mg) were not freeze-dried.

For freeze-dried material	p.value	FDR	Higher in abundance in
18.36_Propanoic acid, 2-methyl-	0.000	0.000	Rectal
21.73_Butanoic acid, 2-methyl-	0.000	0.004	Rectal
29.99_Undecane, 2-methyl-	0.000	0.005	Colon
21.44_Butanoic acid, 3-methyl-	0.000	0.005	Rectal
25.441_Benzaldehyde	0.000	0.007	Rectal
12.55_Acetic acid	0.001	0.009	Rectal
7.19_Propanal	0.001	0.009	Colon
16.07_Propanoic acid	0.001	0.009	Rectal
7.92_Acetic acid, methyl ester	0.001	0.012	Colon
30.91_p-Cresol	0.001	0.012	Rectal
22.85_Nonane, 2-methyl-	0.002	0.020	Colon
30.37_23H-Furanone, 5-ethyldihydro-	0.004	0.029	Rectal
18.25_Hexanal	0.005	0.033	Colon
32.41_2-Decanone	0.006	0.041	Colon
For non-freeze-dried material			
30.91_p-Cresol	0.000	0.031	Rectal
38.91_Tetradecane, 3-methyl-	0.001	0.031	Rectal
39.85_Pentadecane	0.001	0.031	Rectal
19.40_Butanoic acid	0.002	0.031	Rectal
35.85_Tridecane, 2-methyl-	0.003	0.031	Rectal
33.78_Phenol, 4-ethyl-	0.003	0.031	Rectal
32.86_Dodecane, 2-methyl-	0.003	0.031	Rectal
36.89_Tetradecane	0.003	0.031	Rectal
16.50_Toluene	0.004	0.031	Rectal
28.64_Benzeneacetaldehyde	0.004	0.031	Rectal
34.03_Tridecane	0.005	0.031	Rectal
38.70_Heptadecane, 2,6,10,14-tetramethyl-	0.005	0.031	Rectal
16.66_Octane	0.005	0.031	Colon
32.76_Decanal	0.005	0.031	Rectal
36.01_Tridecane, 3-methyl-	0.006	0.033	Rectal
35.46_2-Undecanone	0.009	0.045	Rectal

Table 6.2 VOCs that were significantly different in abundance between the colon and rectal contents (when freeze-dried and not of freeze-dried) 6 horses. Statistical significance was determined using a paired t-test, FDR corrected.

	Mean numbers of VOCs	
Sample	**FD**	**NFD**
Colon	66	44.2[a]
Rectal	74.5	65.2[a]
p-value	0.10	0.002

Shared [a] denotes a significant difference

Table 6.3 The mean number of VOCs in colon and rectal contents (when freeze-dried and not of freeze-dried) of 6 horses. A significantly higher number of VOCs were observed in rectal contents when freeze-drying was performed (paired t-test).

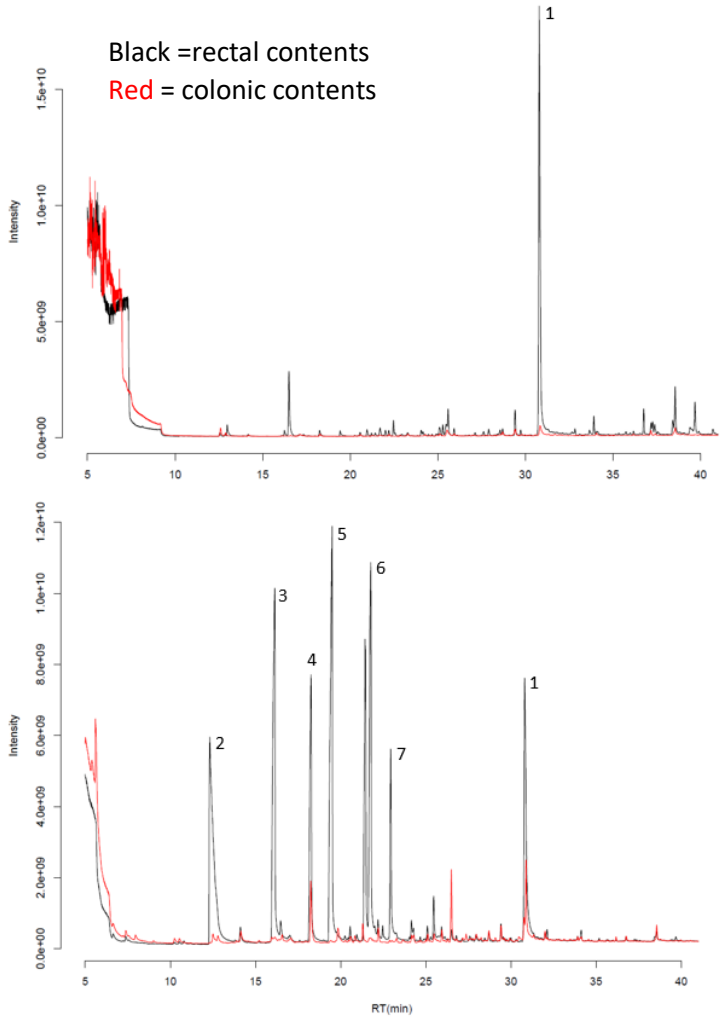


Figure 6.3 Chromatogram overlays of the rectal and colon contents of horse T7. The top chromatogram is an overlay of a non-freeze-dried samples (1000 mg) and the bottom chromatogram is an overlay of freeze-dried samples (100 mg). The black is rectal contents and the red is colon contents. Key: 1 = p-cresol, 2 = acetic acid, 3 = propanoic acid, 4 = propanoic acid, 2-methyl-, 5 = butanoic acid, 6 = butanoic acid, 3-methyl-, 7 = pentanoic acid.

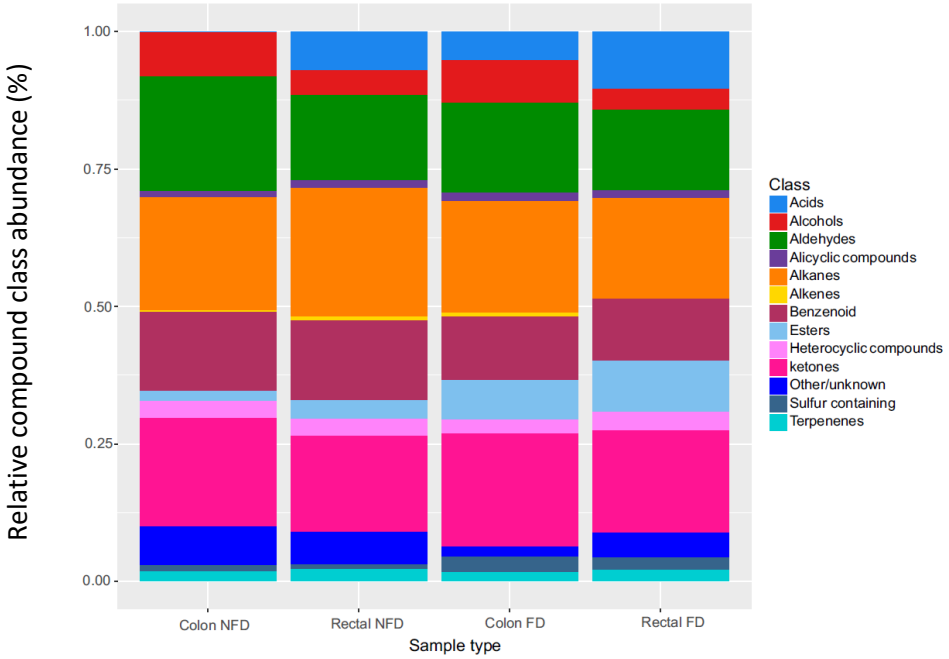


Figure 6.4 A stacked plot of VOCs grouped for chemical class that were found in the colon and rectal contents of 6 horses. Chemical classes were weighted by number of times VOCs belonging to each chemical class were present in samples at the time point displayed on the x-axis. Some key VOCs of various chemical classes were plotted in Appendix 6.1 to show change with respect to volatility. Key: NFD = non-freeze-dried, FD = freeze-dried. FD samples were analysed as a weight of 100 mg and NFD samples were analysed as a weight of 1000 mg.

6.3.3 Comparison of the VOC profiles of AT and CO samples

A PCA plot of AT and CO samples is shown in Figure 6.5. There was no distinct clustering for level of burden or a divide between AT and CO. Three compounds were significantly different between AT and CO. These compounds were: tetradecane, 3-methyl- (p=0.03, t-test); 3-octanone (0.04, t-test) and benzeneacetaldehyde (0.04, t-test). However, post-FDR correction p-values for the three significant compounds were $\geq$0.75.

A sPLS-DA model was built using a subset of VOCs selected by the model. A list of VOCs selected for the sPLS-DA model and included in each component is in Appendix 6.2. The performance of the model is shown in Figure 6.6a and the sPLS-DA plot is in Figure 6.6b. A BER of 45-50% indicated the separation of AT and CO in Figure 6.6b may be over-fitted. The loadings of VOCs important in each group are shown in Figures 6.7a and b for components 1 and 2, respectively.

Using some of the important features from the sPLS-DA model and LASSO, a ROC curve model was built on CO (n=42) and AT (n=22) samples. The following compounds were used to build the model: benzene acetaldehyde; tetradecane, 3-methyl-; 2-butanone; pentanal; 2-pentanone; 1-pentanol; 3-octanone; 1-cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-; cis-2,6-dimethyl-2,6-octadiene; phenol, 4-ethyl-. The ROC curve is shown in Figure 6.8 with an AUC of 0.77, p=0.003 (permutation test). The model based on this training set had specificity and sensitivity values of 86.4% and 81.0%, respectively. The model was tested using a validation subset of 10 CO and 5 AT samples for which the AUROC was 0.84 (Figure 6.9) with an optimum specificity and sensitivity of 70% and 80%, respectively. All tapeworm samples that were misclassified as controls were of low tapeworm burdens (1-20 worms) and therefore below clinical requirements for treatment.

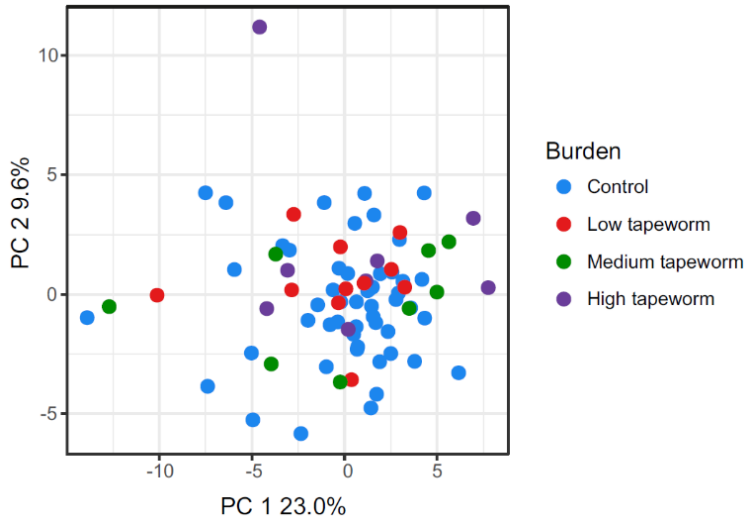
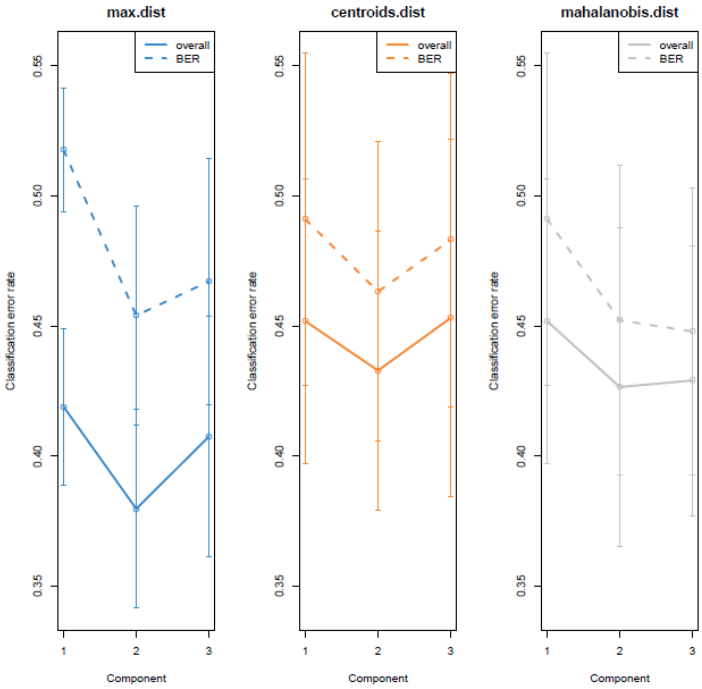


Figure 6.5 A PCA of the VOC profile of rectal contents collected post-mortem from horses that were positive or negative for tapeworm infection. The thresholds for infection were defined as: control (0 tapeworms, n=52), low (1-20, n=11), medium (21-49, n=8) and high (≥50, n=8).

a)



b)

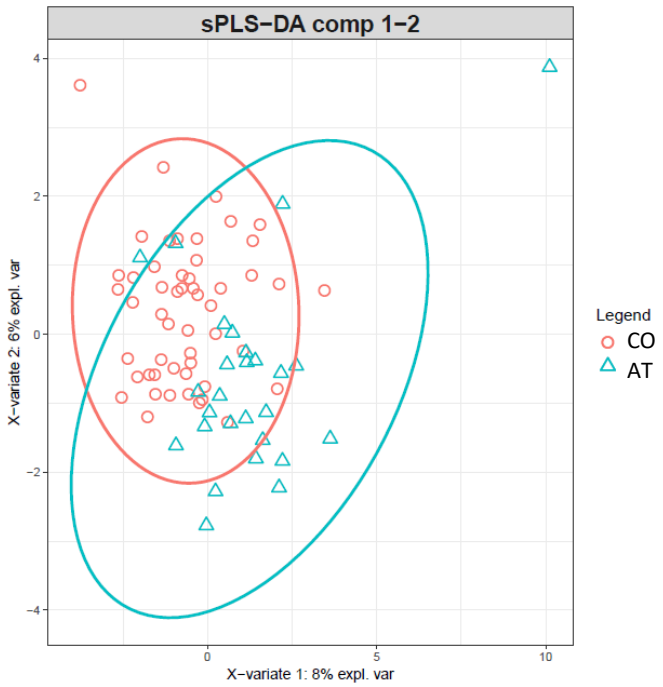


Figure 6.6 In a) the classification error rate of a sPLS-DA model and b) the sPLS-DA plot used to classify horses with and without tapeworm from VOCs detected in rectal contents. Key: ER = error rate, BER = balanced error rate, AT = tapeworm positive, CO = control (tapeworm negative).

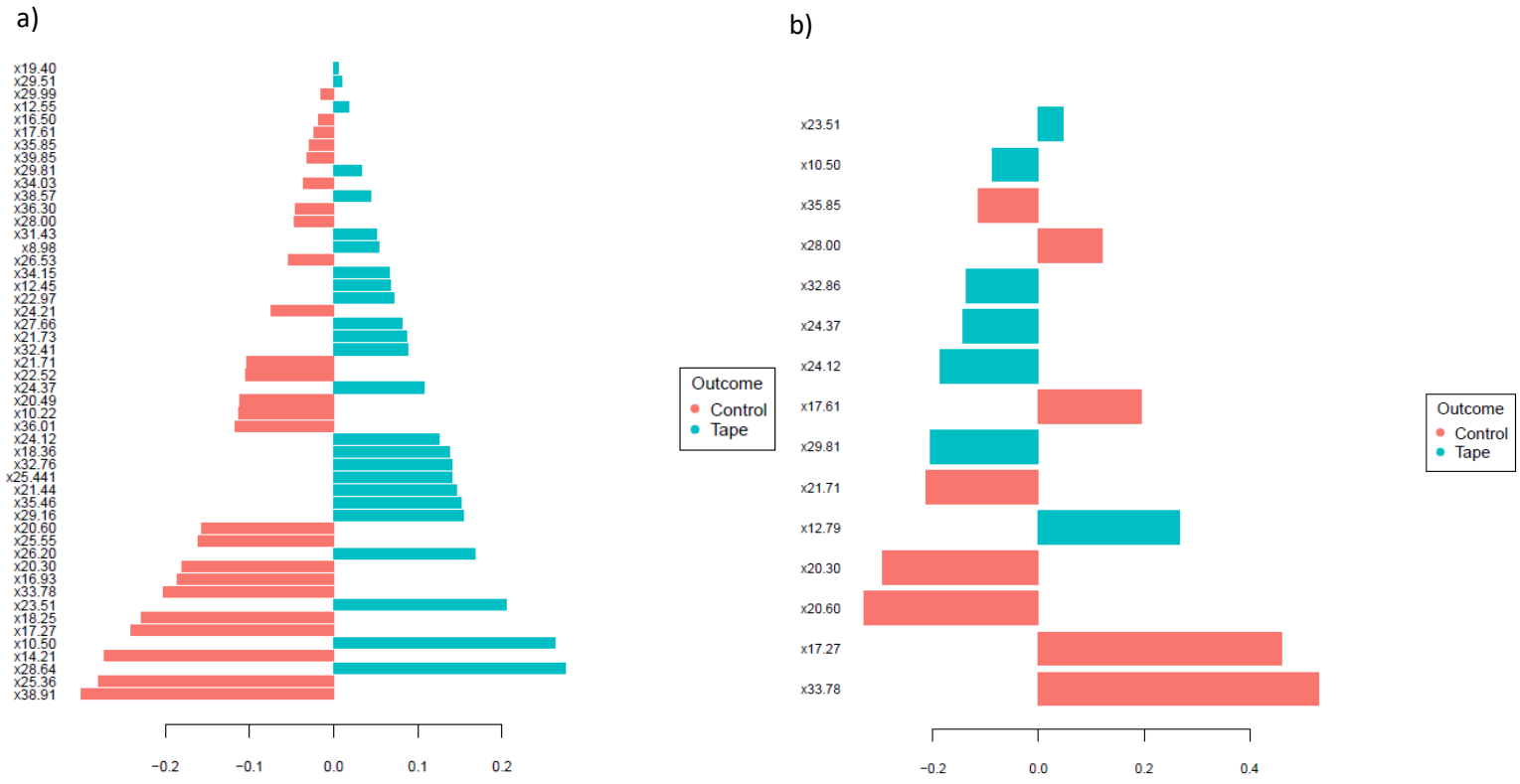


Figure 6.7 sPLS-DA loading scores for the sPLS-DA plot (Figure 6.6b) used to classify horses with and without tapeworm from VOCs detected in rectal contents. In a) scores for component 1 are shown and in b) scores for component 2 are shown. Key: Tape = tapeworm positive, CO = control (tapeworm negative). VOCs are labelled by retention time, names of VOCs can be found in Appendix 6.2.

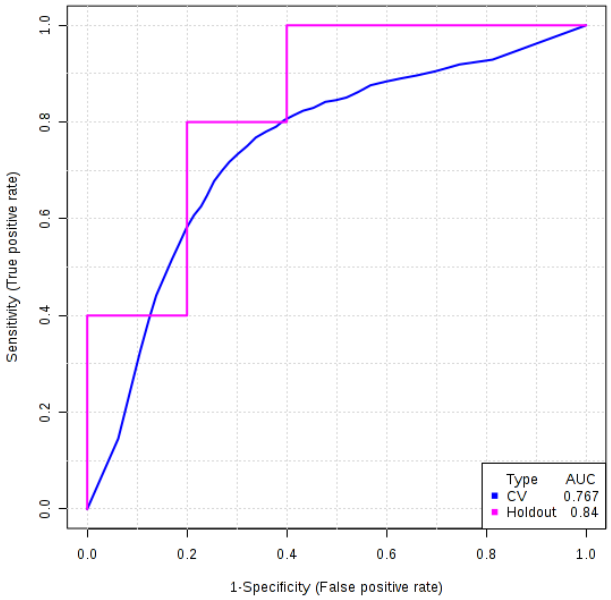


Figure 6.8 ROC curves built using a rectal VOC data to classify horses with and without tapeworm infection. For the training set (AT n=42, CO n=22) an AUC of 0.77 with a 95% confidence interval of 0.56-0.93 was obtained. For the test set (AT n=10, CO n=5) an AUC of 0.84 was calculated. Key AT =horses with ≥1 tapeworm present, CO = control (tapeworm negative).

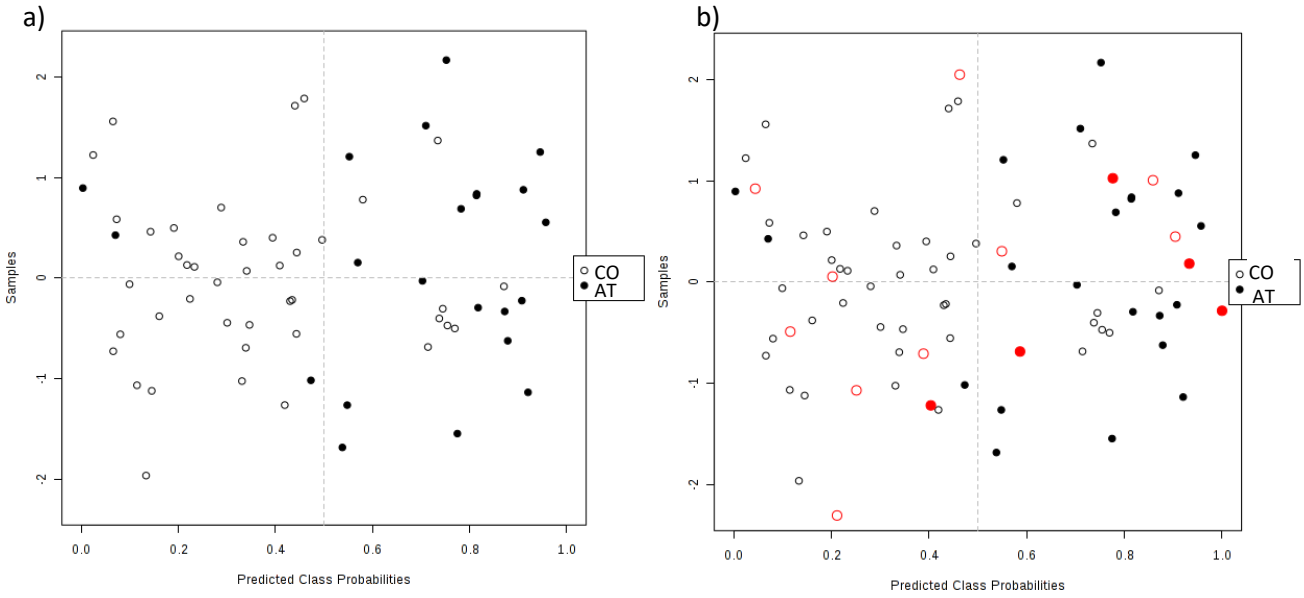


Figure 6.9 The average class probability after 100 cross-validations based on ROC curves built to classify horses with and without tapeworm infection. The cut off for group assignment in is 0.5. In **a)** class probabilities of the training set (CO, n=42 and MH, n=22) are shown and in **b)** the test set (CO, n=10 and MH, n=5) class probabilities are shown. Key MH =horses with ≥21 tapeworms present, CO = control (tapeworm negative).

6.3.4 Comparison of the VOC profiles of MH and CO samples

A PCA plot of the VOC profiles of MH samples and CO is shown in Figure 6.10. No distinct clustering for MH or CO was demonstrated. In a t-test, three compounds were identified as significantly different in abundance between MH and CO. The compounds higher in abundance in CO samples were: 2-pentanone (p=0.01, 0.65), 1-pentanol (p=0.003, 0.24). One compound (decanal, p=0.04) was significantly higher in MH samples. However, post-FDR correction p values were $\geq$0.24.

A sPLS-DA model was built based on VOCs selected by the model. A list of all VOCs included is in Appendix 6.3. The performance of the model is shown in Figure 6.11a and the sPLS-DA plot is in Figure 6.11b. As with AT and CO, a BER of 45-50% indicates the separation of MH and CO in Figure 6.11a may also be over-fitted. The loadings of VOCs which were important in the separation of each group are shown in Figures 6.12a and b for components 1 and 2, respectively.

The same ROC curve model built and tested in section 6.3.4 for AT and CO was applied to MH (n=16) and CO (n=52). A ROC curve with an AUC of 0.76 was generated (Figure 6.13a), p=0.007 (permutation testing) Predicted class probabilities resulted in a specificity of 80.8% and a sensitivity of 87.5% (Figure 6.13b). Further testing demonstrated that the model could not distinguish between control (n=52) and low tapeworm (n=11) burdens (AUC 0.5) or between low tapeworm and MH (AUC 0.5).

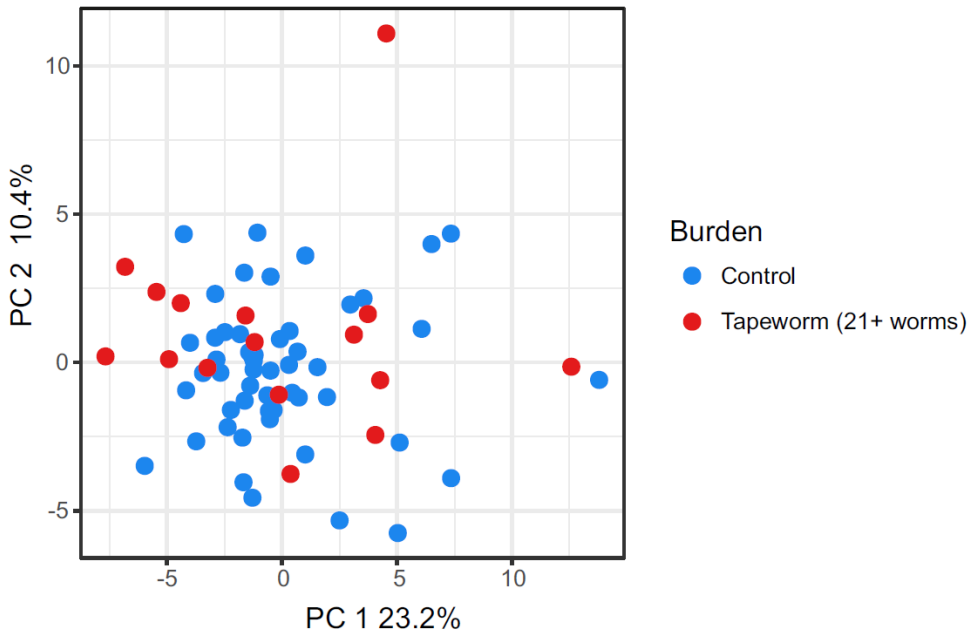


Figure 6.10 A PCA of the VOC profile of rectal contents collected post-mortem from horses that were positive or negative for tapeworm infection.

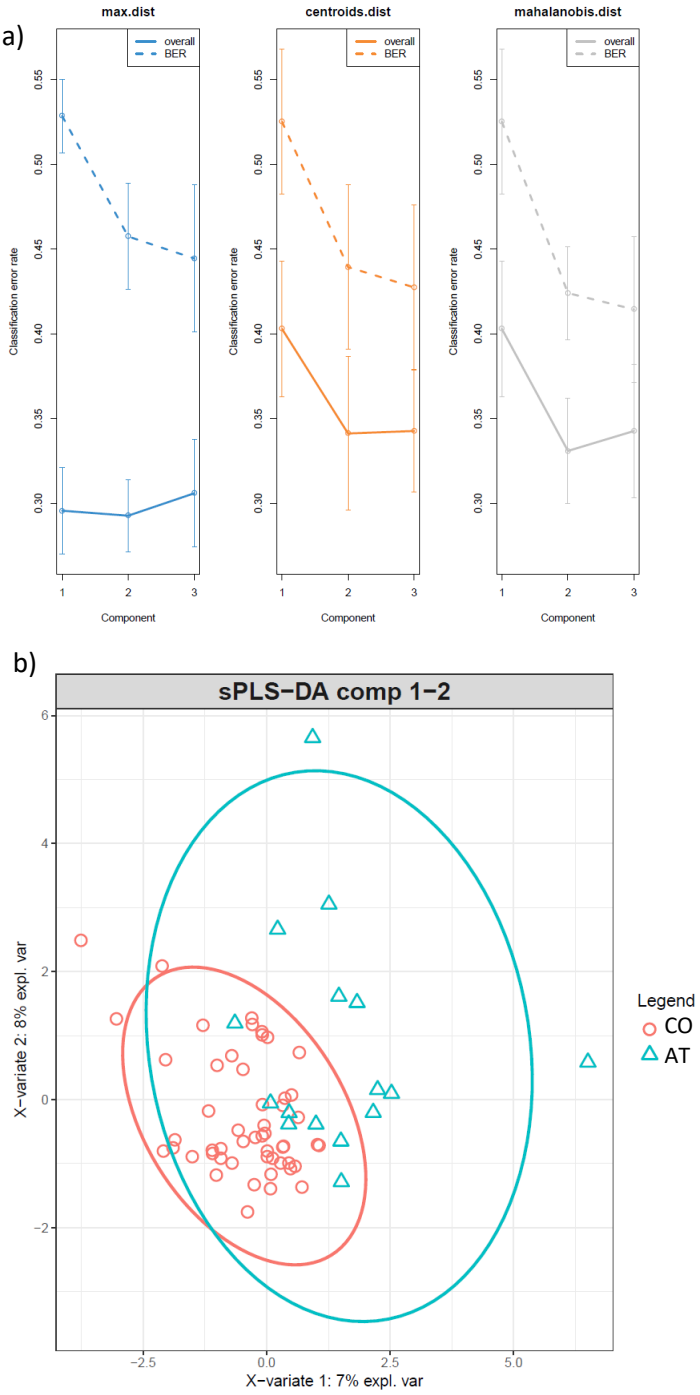


Figure 6.11 In a) the classification error rate of a sPLS-DA model and b) the sPLS-DA plot used to classify horses with and without tapeworm from VOCs detected in rectal contents. Key: ER = error rate, BER = balanced error rate, MH = horses with ≥21 tapeworms present, CO = control (tapeworm negative).

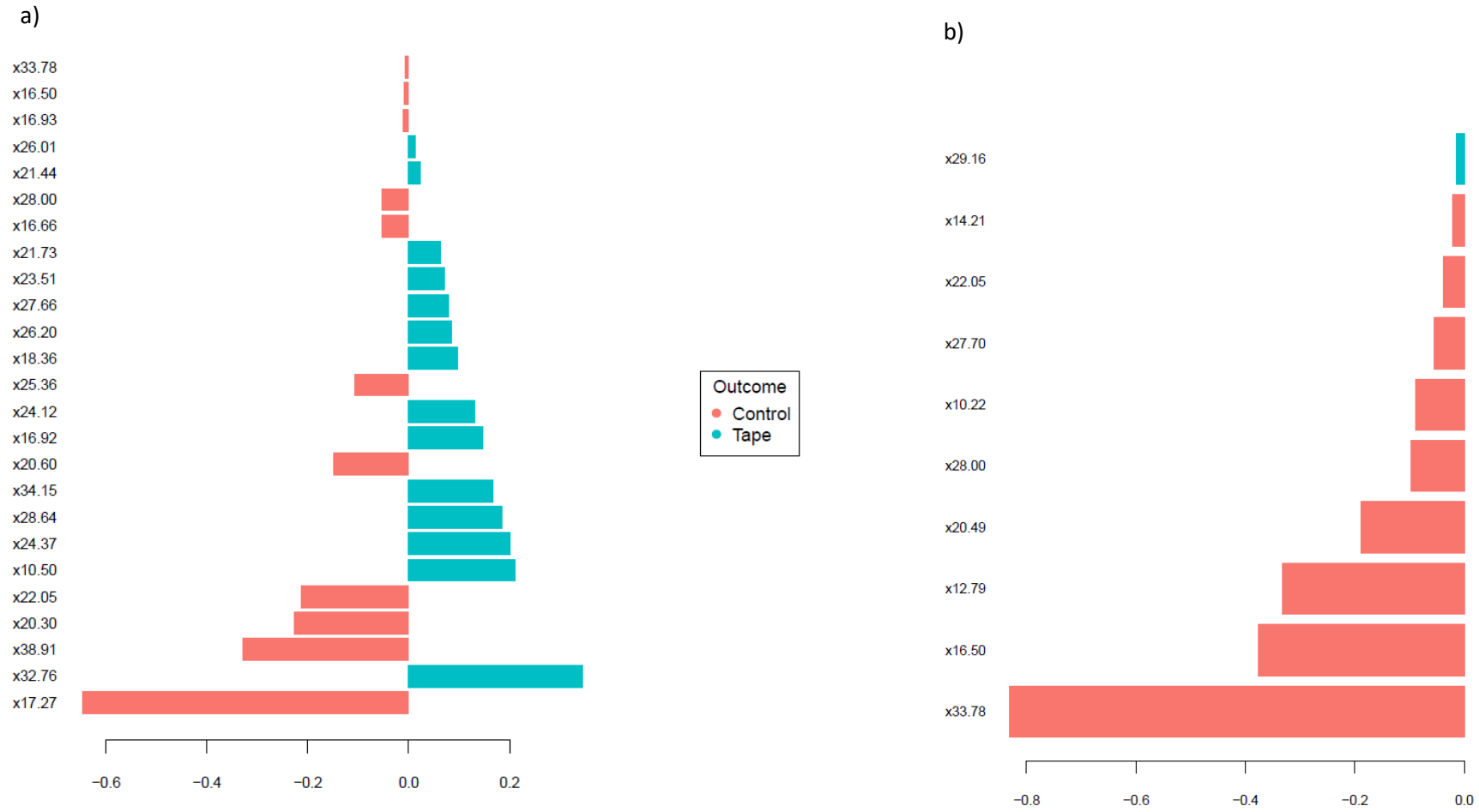


Figure 6.12 sPLS-DA loading scores for the sPLS-DA plot (Figure 6.6b) used to classify horses with and without tapeworm from VOCs detected in rectal contents. In a) scores for component 1 are shown and in b) scores for component 2 are shown. Key: MH = horses with ≥21 tapeworms present, CO = control (tapeworm negative). VOCs are labelled by retention time, names of VOCs can be found in Appendix 6.3.

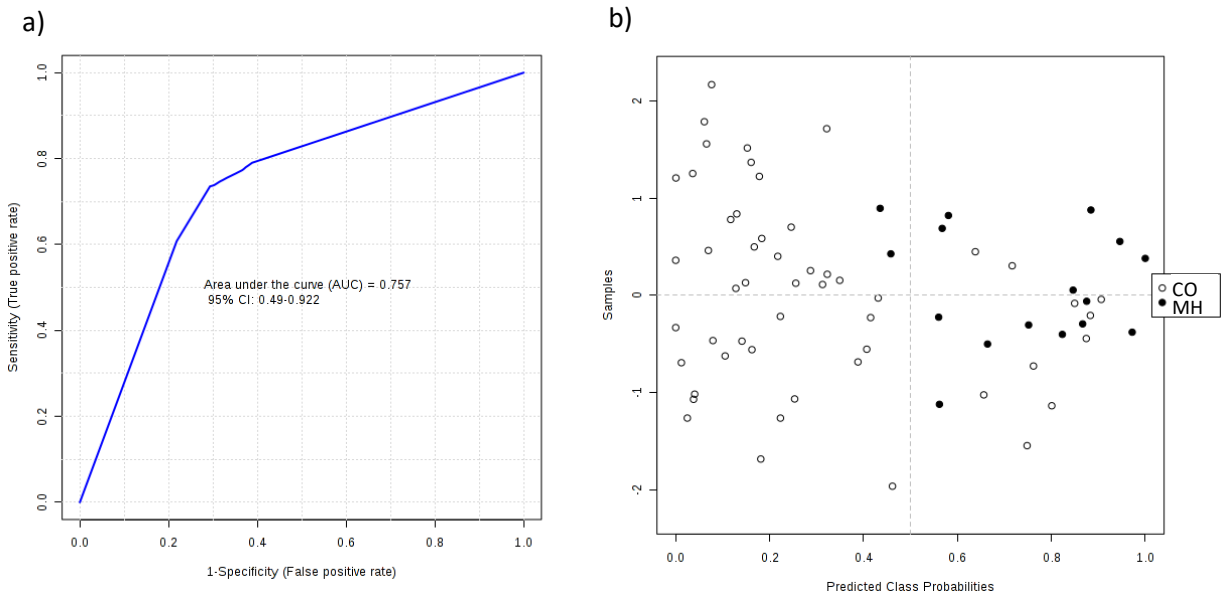


Figure 6.13 A ROC curve in a) and predicted average class probabilities after 100 cross-validations in b) of a rectal VOC model used to classify tapeworm positive (≥21 tapeworms) and tapeworm negative horses. The cut off for group assignment in is 0.5. Key: MH = horses with ≥21 tapeworms present, CO = control (tapeworm negative)

6.3.5 The VOC profile and FEC

PCAs were performed to evaluate the effect of strongyle FEC on VOC profile in horses with and without tapeworm infection. There was no distinct clustering for FEC threshold for low, medium and high groups (Figure 6.14) or when LSC and HSC were compared (Figure 6.15).

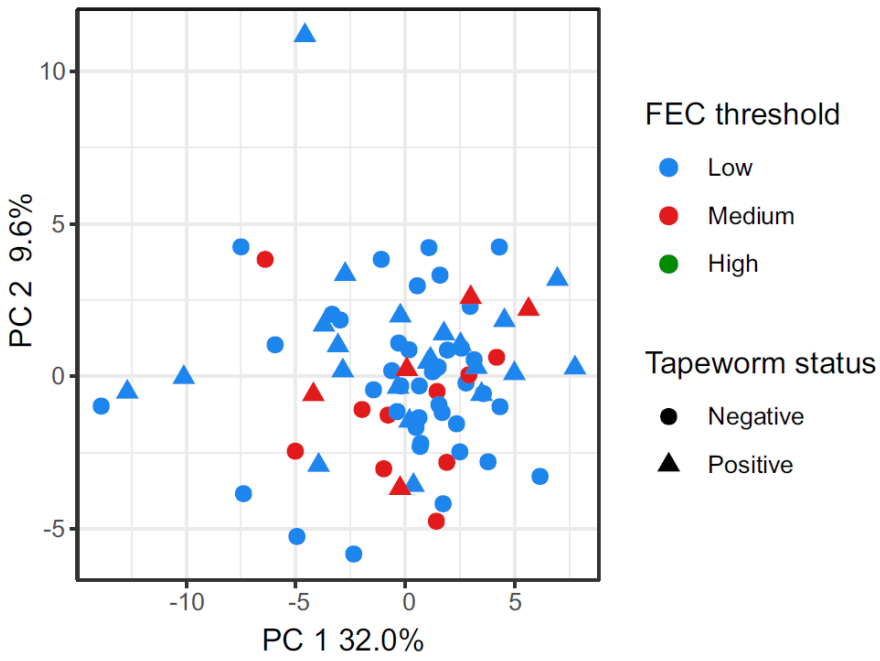


Figure 6.14 A PCA of rectal VOC profiles of horses grouped for strongyle faecal egg count (FEC) and whether samples were positive or negative for tapeworm infection. Strongyle FEC thresholds (e.p.g) were defined as: low (0-199), medium (200-499) and high (≥500).

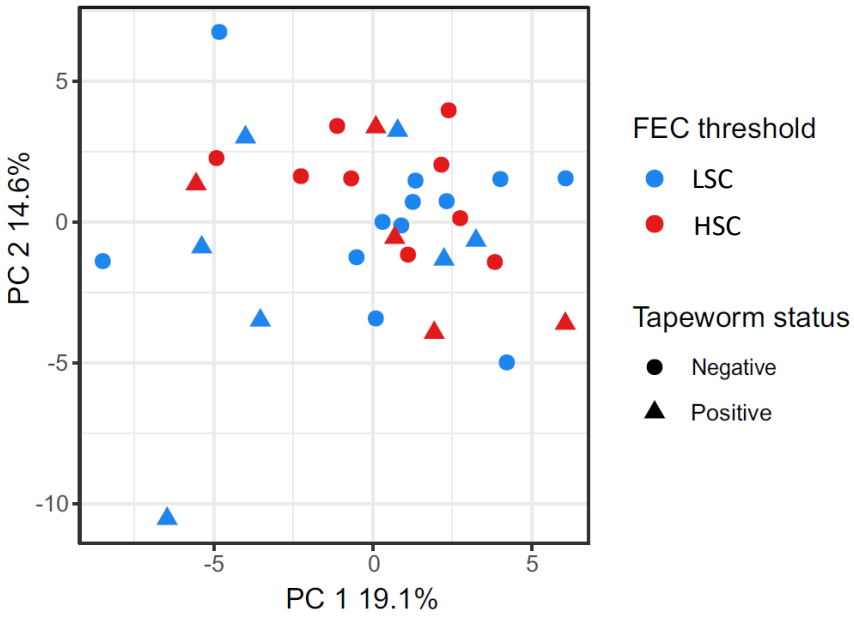


Figure 6.15 A PCA of rectal VOC profiles of horses grouped for strongyle faecal egg count (FEC) and whether samples were positive or negative for tapeworm infection. FEC thresholds were defined as: LSC (0-10 e.p.g) and HSC (≥200 e.p.g.).

6.4 Discussion

6.4.1 Comparison of colonic and rectal contents

A difference in the VOC profile between colon and rectal contents was apparent regardless of whether the dry-weight (freeze-dried) or wet-weight (non-freeze-dried) of samples were compared. The largest differences between colon and rectal contents were observed between NFD samples, possibly because of a higher water content in the colon samples. The sample size (n=6) studied here was small and the diet of the horses was not recorded, therefore, the results should be interpreted with caution. However, the VOC metabolome at different regions of the equine gastrointestinal tract has not been previously investigated. Microbiome studies have shown similarities in bacterial communities in the distal regions of the hindgut, with the caecum showing a very different profile to the rectum and naturally voided faeces (Dougal *et al.*, 2012; Costa *et al.*, 2015a). In microbiome studies, samples have been shown to cluster within horse when distal regions of the intestinal tract are compared. In this VOC study, a pair-wise comparison of colon and rectal contents demonstrated clustering was evident for region rather than horse. Although the pelvic flexure has been shown to be relatively similar to both rectal contents and faeces in terms of bacterial populations, the difference in the VOC metabolome observed here could indicate that the functionality of these bacteria is different in these regions. Dougal *et al.,* (2012) found that VFA concentrations in the caecum, right dorsal colon and faeces were variable. The total VFA as well as acetic acid; propanoic acid, 2-methyl; butanoic acid, 3-methyl; and pentanoic acid, propyl ester concentrations increased from the caecum to the right dorsal colon before dropping again in the rectum, in the latter study. In the current study a number of VFAs were higher in the rectal contents than the pelvic flexure, demonstrating a similar trend to the findings of Dougal *et al.,* (2012). This could indicate that a greater degree of fermentation has occurred by the time the digesta reaches the latter regions of the tract. It would be important to study the VOC metabolome of additional regions including the caecum, ventral and dorsal colons as well as the descending colon to see if faecal VOCs are a reliable proxy for any of these hindgut regions. This would provide further information of the functionality of microorganisms in these regions, which would validate the use of faecal VOCs for monitoring equine gut microorganisms. Furthermore, knowledge of whether faeces are a proxy for hindgut microorganisms would be useful for further studies investigating interactions between intestinal parasites and the gut microbiota in live horses.

6.4.2 Rectal VOCs as biomarkers for tapeworm infection

Unsupervised analysis of the VOC profile did not demonstrate any clustering of samples for tapeworm infection. However, when supervised analysis was performed using a targeted set of VOCs (selected by the sPLS-DA model) some clustering for tapeworm infection was observed. A number of the VOCs important for separating the groups included: decanal, hexanal (both in higher abundance in tapeworm samples) and 1-pentanol which were in higher abundance in control samples. An increase in aldehydes and a decrease in alcohols has previously been characterised in the faeces of humans and mice with intestinal inflammation (Ahmed *et al.*, 2013; Reade *et al.*, 2019). The level of inflammation at the gut mucosa was not recorded in the current study, but increased mucosal inflammation has been commonly reported in horses with ≥21 tapeworms (Pearson *et al.*, 1993; Kjaer *et al.*, 2007; Hreinsdóttir *et al.*, 2019). Increased abundances of decanal and hexanal and a decreased abundance of 1-pentanol was not a signature in the colonic contents of tapeworm-infected horses in **Chapter 5,** despite this region being closer in proximity to where the tapeworms reside in the gut. Comparisons between the rectal contents analysed in this chapter and the colonic contents analysed in **Chapter 5** should be compared loosely as the colonic contents were freeze-dried and the rectal contents in the current chapter were not. However, it may be speculated that as digesta continues along the gastrointestinal tract, further metabolic processes and reactions occur. This may result in end-point metabolites that are different, but representative of earlier events on reaching the rectum. This hypothesis requires further investigation.

A compound included in the ROC curve model (1-cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-) was also found to be more abundant in colonic contents of tapeworm infected horses analysed in **Chapter 5.** As discussed in **Chapter 5,** 1-cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl- significantly correlated with fungal OTUs and may be linked with fungal metabolism in **Chapter 3**. In the current chapter 1-cyclohexene-1-carboxaldehyde and 2,6,6-trimethyl- did not differ in abundance in colonic and rectal samples, implying that faeces could be a potential medium for monitoring this VOC which may have potential as a biomarker for tapeworm infection. The other 9 VOCs included in the model did not mirror the pattern observed in the intestinal contents in **Chapter 5**. The reasons for this may be similar to those described above for the differences in the profiles of decanal, hexanal and 1-pentanol in colon and rectal samples.

The faecal VOC model shared the same limitation as the saliva ELISA in that it was not able to distinguish between control (0) low (1-19 tapeworms for saliva ELISA, 20 for VOC model) and MH (20/21+) burdens (Lightbody *et al.*, 2016). Furthermore, any horses with ≥20 tapeworms were not misdiagnosed by the saliva antibody assay. In the current study all horses with ≥21 tapeworms were correctly classified when the VOC model was applied to AT and CO. Two out of 16 horses were misclassified as controls when MH and CO were included in the VOC model, indicating the VOC model had a marginally poorer performance when tested in another cohort.

Overall, the VOC biomarker model proposed in this chapter did not out-perform current commercially available tests for equine tapeworm infection. The sensitivity and specificity values of training and test data of the faecal VOC model (70-81% sensitivity, 80-86.4% specificity) were similar to those published for the saliva ELISA (83% sensitivity, 85% specificity) (Lightbody *et al.*, 2016). However, the use of a faecal sample rather than saliva has advantages for practical reasons. Data published three years ago reported fewer horse owners used saliva or blood-based tapeworm diagnostic tests than FECs (Easton *et al.*, 2016). If tapeworm and strongyle burdens could be determined from one faecal sample, it may encourage more horse owners to test for tapeworm as well as strongyles.

6.4.3 Limitations and future work

Although a validation set of samples was used to test the model in this study, the reproducibility of the model was not tested. To test the reproducibility of this VOC model a larger cohort would be required. A power calculation based on this pilot data using the function PerformPowerProfiling() in Metaboanalyst, estimated 300 horses per group would provide 90% power (p=0.05). In addition to analysing faecal VOCs, the use of a saliva ELISA on the same horses would also allow a direct comparison between gold standard, the current commercial test and faecal VOCs.

The equine tapeworm species *A. magna* resides in the small intestinal regions of the gastrointestinal tract (Bohórquez *et al.*, 2012). The pathogenicity of *A. magna* is believed to be less than *A. perfoliata* as there have only been a few reports of clinical disease caused by this tapeworm (Bohórquez *et al.*, 2012). Few studies have recorded the prevalence of *A. magna*, but a Spanish abattoir study found 24% and 18% of horses were infected with *A. perfoliata* or *A. magna* only and 11% had mixed infections (Meana *et al.*, 2005). A limitation of both the serological and the saliva ELISA is that neither is able to distinguish between *A. magna* and *A. perfoliata* infections because of cross-reactivity between antigens (Bohórquez

et al., 2012; Lightbody *et al.*, 2016). Therefore, horses that test positive for tapeworm infection by a serological or saliva ELISA may be infected with *A. magna* and may receive an unnecessary dose of anthelmintic. *A. magna* burdens were not determined in this current study; therefore, it is not known whether faecal VOCs could differentiate between *A. magna* and *A. perfoliata* infections. Future abattoir studies should aim to characterise *A. magna* in addition to *A. perfoliata* burdens to determine whether this would be a confounding factor in the faecal VOC model.

Although strongyle FEC did not show clustering based on VOC profile in a PCA analysis, it cannot be ruled out that the presence of strongyles was a confounding factor. Including horses with very low or negative strongyle burdens would have severely restricted the numbers of horses that could have been included in this study, as just 23% of samples had $\leq$10 strongyle e.p.g. Moreover, the limitations of using FEC as a determinant of strongyle burden have been discussed earlier in this thesis.

Determining the origin of VOCs would be an important factor to consider if a diagnostic test would arise from this work. As discussed in **Chapter 1**, VOCs can be universal from multiple sources e.g. the environment, the intestinal microbiota and possibly the parasite itself which would skew results. Going forward the need to exclude exogenous VOCs or identify precise patterns of multiple VOCs associated with tapeworm infection would be needed.

The future for VOCs as a simple and cost-effective diagnostic could emerge in the form of a portable GCMS device (Agbroko & Covington, 2018). The faecal VOC model may be advantageous over current tests because of the practicality in collecting faecal samples instead of saliva. The saliva and serological ELISAs are unsuitable for use in detecting anthelmintic resistance because of slow antibody clearance (Abbott *et al.*, 2008). In future studies it would be useful to investigate at what rate a tapeworm-infected faecal VOC profile would return to a healthy VOC profile after clearance of infection. If faecal VOCs are more informative of real-time infection, it could be a tool for monitoring anthelmintic resistance in tapeworm populations. An additional factor that should be considered in future investigations is how cost-effective a faecal VOC test would compare to currently marketed tests.

Work performed in this thesis has, for the first time, compared the colon and rectal VOC metabolome of the horse using HS-SPME-GCMS. The results demonstrated clustering for gut region rather than horse, indicating a distinct VOC profile of the regions. Further work involving the correlation of microbiome and metabolome data from the latter regions of the

equine gut would allow a better understanding of whether faeces are an accurate indicator of the functional gut microbiome. In this chapter it was also shown that VOCs may have potential for diagnosing tapeworm infection, but validation in larger cohorts is necessary.

Chapter 7 General discussion and conclusions

7.1 General discussion

The first aim of this thesis was to develop a method to extract VOCs from horse faeces. The method was applied to investigate the use of VOCs as a proxy for studying the hindgut microbiome to explore the potential for temporal changes in different cohorts of healthy horses and in the context of finding a biomarker for tapeworm infection. This is an important area of study because changes in the equine gut microbiome can be detrimental to equine health. Furthermore, the interaction of parasites with the gut microbiome is a relatively unexplored area of research in horses. Changes in the metabolome may predict parasite infection and, by association, the risk of particular forms of colic (Clark *et al.*, 2018; Peachey *et al.*, 2018, 2019). Techniques to monitor alterations in the hindgut microbiome that are cost-effective, fast and reliable would be an asset to a veterinary surgeon's toolbox, particularly in horses at increased risk of colic.

A series of method development investigations to extract VOCs from horse faeces were conducted and the optimised method was then applied to different cohorts of horses to begin to address some current gaps in knowledge about the gut microbiota in healthy and parasite-infected horses. Microbiome and metabolome data were integrated to strengthen the evidence of a relationship between the two omics; to date this approach has been applied infrequently in equine studies (Peachey *et al.*, 2019).

A standardised method to extract VOCs from equine faeces for HS-SPME-GCMS had not been published before this research commenced. A method was developed to extract VOCs from equine faeces (**Chapter 2)**, which proved to be different from methods optimised for faeces from other species (human and mouse) (Reade *et al.*, 2014). The method development carried out in this thesis demonstrated the importance of method optimisation for the specific sample matrix to be investigated. In addition, the importance of including method development in the experimental design of metabolomics studies was highlighted.

The longitudinal studies performed (**Chapters 3** and **4)** demonstrated the importance of considering fluctuations in the faecal metabolome over a period of time. In horses grazing at

pasture, with limited management changes, despite observing shifts in the VOC profile and faecal mycobiome, the horses investigated remained healthy. These studies indicated that the healthy faecal VOC metabolome of horses can be highly variable depending on the time of year, and was fairly stable around the time of foaling (parturition). However, the most significant change in the faecal VOC profile of the periparturient mares (**Chapter 4**) was observed when they were introduced to full-time grazing on lush pasture. Although the process of transitioning from a haylage and grass combined diet to a grass only diet was more gradual for the horses sampled over 12 months (**Chapter 3)**, the difference in the VOC profile between the two diets was also an important finding. In both studies, the VOC results mirrored those obtained for the faecal microbiome, indicating use of VOCs as a potential proxy for the microbiome (Salem *et al.*, 2018, 2019), with benefits in lower costs and complexity of statistical analysis. To further strengthen the 'mirroring' observation of VOCs and microorganisms, statistical modelling to integrate VOC and mycobiome data was performed (**Chapter 3)** and integration of VOC and bacterial microbiome data was performed subsequently (**Chapter 5**). In both chapters' strong correlations between microorganisms and VOCs were observed, which gave an indication of the species that were most likely to be active. In contrast to the present study, Peachey *et al.,* (2019) did not detect any significant associations between equine faecal metabolites and 16S rRNA sequence data. In the latter study the authors detected 28 metabolites using NMR spectroscopy from 23 horses sampled at two time points. In this thesis, 84 metabolites were detected from 48 horses with 1,257 significant correlations (0.3 and above) between metabolites and OTUs (**Chapter 5**). Although a similar number of reads were obtained from 16S rRNA sequencing in both studies, the discrepancies in results could be a consequence of the different statistical models used to integrate data, the greater numbers and type of metabolites or a marginally larger sample size of the current study.

Current knowledge of parasite-microbiome interactions in the equine gut is limited (Clark *et al.*, 2018; Peachey *et al.*, 2018). Previously, studies have focused on an association between strongyle FEC and the faecal microbiome (Clark *et al.*, 2018; Peachey *et al.*, 2018, 2019). The latter studies controlled for the diet and other confounding factors (e.g. age, breed etc.) of the live horses that were sampled, which was not possible with the abattoir material collected for this thesis. However, the use of abattoir material allowed tapeworm infection to be accurately determined post-mortem (gold standard) to compare the colonic microbiome and metabolome of horses with known parasite burdens for the first time. The results demonstrated fewer symbiotic species in the tapeworm-infected group. A number of

VOCs in higher abundance in the tapeworm-infected horses (including undecane, 2,6-dimethyl-; octen-3-ol; 2-octanone and furan compounds) were also highly correlated to fungal OTUs in **Chapter 3**. The meaning of these findings for underpinning the physiological mechanisms of parasite-associated colic was beyond the scope of this study. The observation of fewer symbiotic bacteria in the tapeworm-infected samples agreed with a previous study in horses with high strongyle FECs, and warrants further research (Clark *et al.*, 2018).

Current diagnostic tests for tapeworm infection utilise saliva or serum and are based on immune-response to the A. perfoliata. These have a lag phase, which limits accurate prediction of current infection status, particularly in horses with high infections following anti-cestode treatment. The use of a diagnostic test to accurately determine tapeworm infection in live horses would facilitate studies investigating the microbiome and mycobiome of infected and non-infected horses. In this thesis a VOC biomarker model consisting of 10 VOCs was able to correctly identify 80% of horses with tapeworm infection when tested. One of the VOCs included in the model (1-cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-) was associated with bacterial OTUs (family Prevotellaceae) in colon contents analysed in **Chapter 5** and was associated with a number of fungal OTUs in **Chapter 3**. These correlations suggest 1-cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl- may act as a potential marker for these microorganisms in the faeces and in turn could also be associated with tapeworm infection. These results warrant further investigation of the role of bacteria and fungi in tapeworm infection as well as VOCs as biomarkers. The majority of VOCs in the faecal biomarker model did not correlate with bacterial or fungal OTUs in the previous chapters. This could suggest that these VOCs are related to a host response to parasite infection, the parasite itself or other unidentified microorganisms. However, it was difficult to make direct comparisons between the colon contents analysed in **Chapter 5** and rectal contents analysed in **Chapter 6** for the following reasons: one set of samples was freeze-dried and the other not, additional samples were added (and some removed) to the cohort in **Chapter 6** and the use of a batch effect correction tool meant the data were processed in a different way to **Chapter 5**.

7.2 Overall limitations

Limitations specific to each study were discussed in earlier chapters, but overall limitations of this thesis are discussed in this section. Investigation of the equine faecal VOC metabolome is a relatively recent and under-studied area, with few published studies on which the sample size / power calculation could be based (Turner *et al.*, 2013; Escalona *et al.*, 2014; Proudman *et al.*, 2015). This made the choice of sample size for the investigations in this thesis difficult

and potentially they were under-powered. Therefore, much of the data generated in this thesis should be considered as a pilot and used to determine sample sizes for future work.

A small number of studies have demonstrated a change in faecal VOCs following the inclusion of dietary supplements to horses (Proudman *et al.*, 2015; Snalune *et al.*, 2019). In agreement with these studies, investigations in this thesis provide additional evidence that faecal VOCs can alter according to dietary change. In the longitudinal studies, changes in diet were recorded, but it could not be concluded when the VOC alterations occurred and whether other confounding factors, e.g. management changes, breed of horse, parasite burden and weather changes were also involved. A crossover design study to include a wash-out period between diets would be needed to fully understand the extent to which faecal VOCs are affected by diet. Furthermore, the longitudinal analysis demonstrated that one time-point investigations including the method development investigations in **Chapter 2** and the tapeworm studies in **Chapters 5** and **6** should be interpreted carefully as the healthy faecal VOC metabolome has been shown to be variable over time.

A broad limitation of this thesis was that just one metabolomics platform (GCMS) and extraction method (SPME) was used. To fully encompass the metabolome and perform detailed pathway analysis the use of multiple platforms (e.g. LCMS and NMR) would provide a much more in-depth analysis of the equine hindgut metabolome (de Raad *et al.*, 2016b). A further limitation of the general methods used in this thesis was the application of a database to putatively identify VOCs. Three standards were analysed in the same way as samples (2-pentanone, benzaldehyde, and indole) which worked towards MSI guidelines level 1 (Sumner *et al.*, 2007). However, to fully validate the identity of each VOC a library of standards would be required. This was not practical and too expensive for the pilot studies in this thesis but future studies should consider validation of compounds using standards.

7.3 Future work and implications of this thesis

Although disruption to the hindgut microbiome had been linked with colic (Daly *et al.*, 2012), VOCs were not studied in context of colic in this thesis. A pilot study observed a different VOC profile in horses that were presented at a veterinary hospital for SCOD colic rather than non-gastrointestinal related disorders (Turner *et al.*, 2013). A change in the faecal microbiota of periparturient mares prior to a colic episode has also been reported (Weese *et al.*, 2014). Microbiome studies are too labour intensive and expensive for routine monitoring of gut microorganisms. The findings of this thesis warrant studies into VOC biomarkers for monitoring changes in the gut microorganisms preceding a colic episode. Therefore, studying

faecal VOCs and the microbiome longitudinally in a cohort of horses before, during and after a colic episode would provide further information of whether faecal VOCs would have the potential to act as a monitoring tool for colic in horses. The use of a colic monitoring tool would be most effective when horses are known to be at increased risk of colic e.g. have had recurrent colic episodes, broodmares post foaling or acute changes in management such as altered exercise and feeding in horses on box-rest (Archer & Proudman, 2006). Pre-colic sampling may indicate how long changes in the faeces occur prior to a colic episode. Post colic samples may indicate how long after an episode the faeces return to a healthy state, which has implications for recurrence and recovery. The application of the mixOmics R package (Rohart *et al.*, 2017) applied to microbiome and metabolome data would strengthen hypotheses that the microbiota plays a role in colic. The proposed experimental design could lead to the development of a monitoring tool as well as evidence to warrant further investigation of prevention strategies for colic involving manipulation of the microbiome.

An increased abundance of fungal compounds in the autumn of the longitudinal study and in the tapeworm-infected horses was observed in this thesis. The relevance of this finding in terms of tapeworm infection requires further investigation. A good starting point would be a repeat of the abattoir study at multiple times of year to determine whether the same pattern of fungal VOCs is observed all year round in horses with tapeworm burdens or if it is a pattern associated with the autumn only. Taking mucosa samples from inflamed and non-inflamed regions may also be useful when investigating the involvement of the microbiome and mycobiome in parasite driven-inflammation. Establishing a potential role of the gut microbiota in exacerbating parasite-driven inflammation may lead to treatments that involve the manipulation of the microbiome (e.g. probiotics) as an alternative to anthelmintics. The faecal VOC biomarker proposed in this thesis was very much a pilot study because of the small sample size. As mentioned in the discussion section in **Chapter 6** a large cohort would be needed to test the reproducibility of the model.

7.4 Final conclusions

A strength of this thesis was the inclusion of longitudinal studies to characterise the healthy VOC metabolome in two cohorts of horses. Knowledge of the wide fluctuations of the healthy VOC metabolome was essential for interpreting the comparisons between VOC metabolome of tapeworm-infected and non-infected horses. Furthermore, the longitudinal studies will provide a reference for future equine hindgut metabolome studies investigating equine health and disease. Studies of the temporal faecal mycobiome of the horse have not been

published. Here, it was identified that a potentially inverse relationship between anaerobic and facultative-anaerobic fungi may exist. A further finding was an increase in fungal diversity which was a characteristic of grazing horses in the autumn. The relevance of these findings in terms of equine health requires further investigation and this thesis has provided groundwork for future study in this area. Strong positive and negative correlations of VOCs with bacteria and fungi were identified by statistically integrating microbiome and VOC data. The integrated approach provided evidence for potentially active species, as well as strengthening the hypothesis that VOCs could act as a proxy for the microbiome. This thesis has also contributed to the growing area of research investigating the interactions between parasites and the microbiome in horses, which merits further research to develop better therapies and biomarkers in this area.

References

Abbott, J.B., Mellor, D.J., Barrett, E.J., Proudman, C.J. & Love, S. (2008) Serological changes observed in horses infected with Anoplocephala perfoliata after treatment with praziquantel and natural reinfection. *Veterinary Record*, **162**, 50–53.

Abrão, F.O., Duarte, E.R., Freitas, C.E.S., Vieira, E.A., Geraseev, L.C., Silva-Hughes, A.F. da, *et al.* (2014) Characterization of fungi from ruminal fluid of beef cattle with different ages and raised in tropical lignified pastures. *Current Microbiology*, **69**, 649–659.

Achá, S.J., Kühn, I., Mbazima, G., Colque-Navarro, P. & Möllby, R. (2005) Changes of viability and composition of the Escherichia coli flora in faecal samples during long time storage. *Journal of Microbiological Methods*, **63**, 229–238.

Agbroko, S.O. & Covington, J. (2018) A novel, low-cost, portable PID sensor for the detection of volatile organic compounds. *Sensors and Actuators, B: Chemical*, **275**, 10–15.

Aggio, R., Villas-Bôas, S.G. & Ruggiero, K. (2011) Metab: an R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics (Oxford, England)*, **27**, 2316–8.

Aggio, R.B.M., Mayor, A., Coyle, S., Reade, S., Khalid, T., Ratcliffe, N.M., *et al.* (2016) Freeze-drying: an alternative method for the analysis of volatile organic compounds in the headspace of urine samples using solid phase micro-extraction coupled to gas chromatography - mass spectrometry. *Chemistry Central Journal*, **10**, 9.

Agneessens, J., Debever, P., Engelen, S. & Vercruysse, J. (1998) The prevalence of Anoplocephala perfoliata in horses in Belgium, and evaluation of a diagnostic sedimentation/flotation technique. *Vlaams Diergeneeskundig Tijdschrift*, **67**, 27–31.

Ahmed, I., Greenwood, R., Costello, B., Ratcliffe, N. & Probert, C.S. (2016) Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. *Alimentary Pharmacology and Therapeutics*, **43**, 596–611.

Ahmed, I., Greenwood, R., Costello, B.L. de, Ratcliffe, N.M. & Probert, C.S. (2013) An Investigation of Fecal Volatile Organic Metabolites in Irritable Bowel Syndrome. *PLoS ONE*, **8**, 1–13.

Akin, D. & Rigsby, L. (1985) Influence of phenolic acids on rumen fungi. *Agronomy Journal*, **77**, 180–182.

Almeida, M.L.M. De, Feringer, W.H., Carvalho, J.R.G., Rodrigues, I.M., Jordão, L.R., Fonseca, M.G., *et al.* (2016) Intense exercise and aerobic conditioning associated with chromium or L-carnitine supplementation modified the fecal microbiota of fillies. *PLoS ONE*, **11**, 1–21.

Amann, A., Costello, B.D.L., Miekisch, W., Schubert, J., Buszewski, B., Pleil, J., *et al.* (2014) The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *Journal of Breath Research*, **8**, 034001.

Amrane, S., Raoult, D. & Lagier, J.-C. (2018) Metagenomics, culturomics and the human gut microbiota. *Expert Review of Anti-infective Therapy*, **16**, 14787210.2018.1467268.

Archer, D.C., Pinchbeck, G.L., Proudman, C.J. & Clough, H.E. (2006) Is equine colic seasonal? Novel application of a model based approach. *BMC veterinary research*, **2**, 27.

Archer, D.C. & Proudman, C.J. (2006) Epidemiological clues to preventing colic. *The Veterinary Journal*, **172**, 29–39.

Argenzio, R.A. & Hintz, H.F. (1970) Glucose tolerance and effect of volatile fatty acid on plasma glucose concentrations in ponies. *Journal of Animal Science*, **30**, 514–518.

Argenzio, R.A., Lowe, J.E., Pickard, D.W. & Stevens, C.E. (1974a) Digesta equine passage and water exchange in the large intestine. *American Journal of Physiology*, **226**, 1035–1042.

Argenzio, R.A., Southworth, M. & Stevens, C.E. (1974b) Sites of organic acid production and absorption in the equine gastrointestinal tract. *American Journal of Physiology*, **226**, 1043–1050.

Arrazuria, R., Elguezabal, N., Juste, R.A., Derakhshani, H. & Khafipour, E. (2016) Mycobacterium avium subspecies paratuberculosis infection modifies gut microbiota under different dietary conditions in a rabbit model. *Frontiers in Microbiology*, **7**, 1–14.

Austin, D.B. (2019) EquiSal user manual [WWW Document]. URL http://equisal.co.uk/epages/d7497350-0c56-4e11-a207-b31ca0b55b54.sf/en_GB/?ObjectPath=/Shops/d7497350-0c56-4e11-a207-b31ca0b55b54/Categories/The_EquiSal_Test/How_to_collect_a_saliva_sample [accessed on .

Bader, T., Schulz, W., Kümmerer, K. & Winzenbacher, R. (2016) General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry. *Analytica Chimica Acta*, **935**, 173–186.

Bates, J., Mittage, E., Kuhlman, J., Baden, K., Cheeseman, S. & Guillemin, K. (2006) Distinct signals from the microbiota promote different aspects of zebrafish gut differenciation. *Developmental Biology*, **297**, 374–386.

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R. & Fierer, N. (2011) Examining the global distribution of dominant archaeal populations in soil. *ISME Journal*, **5**, 908–917.

Baverud, V., Gustafsson, A., Franklin, A., Lindholm, A. & Gunnarsson, A. (1997) Clostridium difficile associated with acute colitis in mature horses treated with antibiotics. *Equine Veterinary Journal*, **29**, 279–284.

Becker, E. (1932) The present status of problems relating to the ciliates of ruminants and Equidae. *The Quarterly Review of Biology*, **7**, 282–297.

Beckers, K.F., Schulz, C.J. & Childers, G.W. (2017) Rapid regrowth and detection of microbial contaminants in equine fecal microbiome samples. *PLoS ONE*, **12**, 1–18.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57**, 289–300.

Berg, M. Van den, Hoskin, S.O., Rogers, C.W. & Grinberg, A. (2013) Fecal pH and Microbial Populations in Thoroughbred Horses During Transition from Pasture to Concentrate Feeding. *Journal of Equine Veterinary Science*, **33**, 215–222.

Bergman, E.N. (1990) Energy Contributions of Volatile Fatty Acids From the Gastrointestinal Tract in Various Species. *Physiological reviews*, **70**, 567–590.

Bergmann, A., Trefz, P., Fischer, S., Klepik, K., Walter, G., Steffens, M., *et al.* (2015) In vivo volatile organic compound signatures of Mycobacterium avium subsp. paratuberculosis. *PLoS ONE*, **10**, 1–20.

Berkhout, D., Benninga, M., Stein, R. van, Brinkman, P., Niemarkt, H., Boer, N. de, *et al.*

(2016) Effects of Sampling Conditions and Environmental Factors on Fecal Volatile Organic Compound Analysis by an Electronic Nose Device. *Sensors*, **16**, 1967.

Bernatchez, E., Gold, M.J., Langlois, A., Blais-lecours, P., Boucher, M., Duchaine, C., *et al.* (2017) Methanosphaera stadtmanae induces a type IV hypersensitivity response in a mouse model of airway inflammation. *Physiological Reports*, **5**, e13163.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L. & Trees, E. (2018) Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical Microbiology and Infection*, **24**, 335–341.

Biddle, A.S., Black, S.J. & Blanchard, J.L. (2013) An in vitro model of the horse gut microbiome enables identification of lactate-utilizing bacteria that differentially respond to starch induction. *PloS one*, **8**, e77599.

Bijlsma, S., Bobeldijk, I., Verheij, E.R., Ramaker, R., Kochhar, S., Macdonald, I.A., *et al.* (2006) Large-scale human metabolomics studies: A strategy for data (pre-) processing and validation. *Analytical Chemistry*, **78**, 567–574.

Biswal, D., Nandi, A.P. & Chatterjee, S. (2016) Helminth–bacteria interaction in the gut of domestic pigeon Columba livia domestica. *Journal of Parasitic Diseases*, **40**, 116–123.

Blackmore, T.M., Dugdale, A., Argo, C.M., Curtis, G., Pinloche, E., Harris, P. a, *et al.* (2013) Strong stability and host specific bacterial community in faeces of ponies. *PloS one*, **8**, e75079.

Bohórquez, A., Meana, A. & Luzón, M. (2012) Differential diagnosis of equine cestodosis based on E/S and somatic Anoplocephala perfoliata and Anoplocephala magna antigens. *Veterinary Parasitology*, **190**, 87–94.

Bohórquez, G.A., Luzón, M., Martín-Hernández, R. & Meana, A. (2015) New multiplex PCR method for the simultaneous diagnosis of the three known species of equine tapeworm. *Veterinary parasitology*, **207**, 56–63.

Bond, A., Vernon, A., Reade, S., Mayor, A., Minetti, C., Wastling, J., *et al.* (2015) Investigation of Volatile Organic Compounds Emitted from Faeces for the Diagnosis of Giardiasis. *Journal of Gastroenterology and Liver Disease*, **24**, 281–286.

Boswinkel, M. & Sloet van Oldruitenborgh-Oosterbann, M. (2007) Correlation between colic and antibody levels against Anoplocephala perfoliata in horses in The Netherlands. *Tijdschrift Voor Diergeneeskunde*, **132**, 508–512.

Braga, F.R., Araújo, J.V., Silva, A.R., Araujo, J.M., Carvalho, R.O., Tavela, A.O., *et al.* (2009) Biological control of horse cyathostomin (Nematoda: Cyathostominae) using the nematophagous fungus Duddingtonia flagrans in tropical southeastern Brazil. *Veterinary Parasitology*, **163**, 335–340.

Brosschot, T.P. & Reynolds, L.A. (2018) The impact of a helminth-modified microbiome on host immunity. *Mucosal Immunology*, **11**, 1039–1046.

Burnham, S., Zierer, J., Jackson, M.A., Kastenmüller, G., Mangino, M., Long, T., *et al.* (2018) The fecal metabolome as a functional readout of the gut microbiome. *Nature Genetics*, **50**, 790–795.

Cai, H., Archambault, M. & Prescott, J.F. (2003) 16S ribosomal RNA sequence – based identification of veterinary clinical bacteria. *Journal of Veterinary Diagnostic Investigation*, **9**, 465–469.

Cai, L., Koziel, J.A., Davis, J., Lo, Y.C. & Xin, H. (2006) Characterization of volatile organic compounds and odors by in-vivo sampling of beef cattle rumen gas, by solid-phase microextraction, and gas chromatography-mass spectrometry-olfactometry. *Analytical and Bioanalytical Chemistry*, **386**, 1791–1802.

Cameron, S.J.S. & Takáts, Z. (2018) Mass spectrometry approaches to metabolic profiling of microbial communities within the human gastrointestinal tract. *Methods*, **149**, 13–24.

Caporaso, J.G., Bittinger, K., Bushman, F.D., Desantis, T.Z., Andersen, G.L. & Knight, R. (2010a) PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics*, **26**, 266–267.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al.* (2010b) QIIME allows analysis of high- throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Cattadori, I.M., Sebastian, A., Hao, H., Katani, R., Albert, I., Eilertson, K.E., *et al.* (2016) Impact of helminth infections and nutritional constraints on the small intestine microbiota. *PLoS ONE*, **11**, 1–23.

Cebra, J.J. (1999) Influences of microbiota on intestinal immune system development. *American Journal of Clinical Nutrition*, **69**, 1046s–1051s.

Chapman, A. (2009) Acute diarrhea in hospitalized horses. *Veterinary Clinics of North America: Equine Practice*, **25**, 363–380.

Chapman, M., French, D. & Klei, T. (2002) Gastrointestinal helminths of ponies in Louisiana: a comparison of species currently prevalent with those present 20 years ago. *Journal of Parasitology*, **88**, 1130–1134.

Chappell, C.L., Darkoh, C., Shimmin, L., Farhana, N., Kim, D.-K., Okhuysen, P.C., *et al.* (2016) Fecal indole as a biomarher of susceptibility to Cryptosporidium infection. *Infection and Immunity*, **84**, IAI.00336-16.

Cheng, Y.F., Jin, W., Mao, S.Y. & Zhu, W.Y. (2013) Roduction of citrate by anaerobic fungi in the presence of co-culture methanogens as revealed by 1H NMR spectrometry. *Asian-Australasian Journal of Animal Sciences*, **26**, 1416–1423.

Cho, D.H., Kong, S.H. & Oh, S.G. (2003) Analysis of trihalomethanes in drinking water using headspace-SPME technique with gas chromatography. *Water Research*, **37**, 402–408.

Christie, M. & Jackson, F. (1982) Specific identification of strongyle eggs in small samples of sheep faeces. *Research in Veterinary Science*, **32**, 113–117.

Cichorska, B., Komosa, M., Nogowsk, L., Maćkowiak, P. & Józefia, D. (2014) Significance of Nutrient Digestibility in Horse Nutrition – A Review. *Annals of Animal Science*, **14**, 779–797.

Clark, A., Sallé, G., Ballan, V., Reigner, F., Meynadier, A., Cortet, J., *et al.* (2018) Strongyle infection and gut microbiota: Profiling of resistant and susceptible horses over a grazing season. *Frontiers in Physiology*, **9**, doi: 10.3389/fphys.2018.00272.

Clement, B.G., Kehl, L.E., DeBord, K.L. & Kitts, C.L. (1998) Terminal restriction fragment patterns (TRFPs), a rapid, PCR- based method for the comparison of complex bacterial communities. *Journal of Microbiological Methods*, **31**, 135–142.

Combet, E., Henderson, J., Eastwood, D.C. & Burton, K.S. (2006) Eight-carbon volatiles in mushrooms and fungi: Properties, analysis, and biosynthesis. *Mycoscience*, **47**, 317–326.

Costa, M.C., Arroyo, L.G., Allen-Vercoe, E., Stämpfli, H.R., Kim, P.T., Sturgeon, A., *et al.* (2012) Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput sequencing of the V3-V5 region of the 16S rRNA gene. *PloS one*, **7**, e41484.

Costa, M.C., Silva, G., Ramos, R.V., Staempfli, H.R., Arroyo, L.G., Kim, P., *et al.* (2015a) Characterization and comparison of the bacterial microbiota in different gastrointestinal tract compartments in horses. *The Veterinary Journal*, **205**, 74–80.

Costa, M.C., Stämpfli, H.R., Allen-Vercoe, E. & Weese, J.S. (2016) Development of the faecal microbiota in foals. *Equine veterinary journal*, **48**, 681–688.

Costa, M.C., Stämpfli, H.R., Arroyo, L.G., Allen-Vercoe, E., Gomes, R.G. & Weese, J.S. (2015b) Changes in the equine fecal microbiota associated with the use of systemic antimicrobial drugs. *BMC Veterinary Research*, **11**, 1–12.

Costa, M.C. & Weese, J.S. (2012) The equine intestinal microbiome. *Animal health research reviews / Conference of Research Workers in Animal Diseases*, **13**, 121–8.

Cotter, P., Stanton, C., Ross, R. & Hill, C. (2012) The impact of antibiotics on the gut microbiota as revealed by high throughput DNA sequencing. *Discovery Medicine*, **13**, 193–199.

Couch, R.D., Navarro, K., Sikaroodi, M., Gillevet, P., Forsyth, C.B., Mutlu, E., *et al.* (2013) The approach to sample acquisition and its impact on the derived human fecal microbiome and VOC metabolome. *PloS one*, **8**, e81163.

Cripps, A. & Williams, V. (1975) The effect of pregnancy and lactation on food intake, gastrointestinal anatomy and the absorptive capacity of the small intestine in the albino rat. *British Journal of Nutrition*, **33**, 17–32.

Cummings, J.H. & Macfarlane, G.T. (1991) The control and consequences of bacterial fermentation in the human colon. *Journal of Applied Bacteriology*, **70**, 443–459.

Curtis, L., Burford, J., England, G.C.W. & Freeman, S. (2019) Risk factors for acute abdominal pain (colic) in the adult horse : A scoping review of risk factors , and a systematic review of the effect of management-related changes. *PLoS ONE*, **14**, e0219307.

Daly, K., Proudman, C.J., Duncan, S.H., Flint, H.J., Dyer, J. & Shirazi-Beechey, S.P. (2012) Alterations in microbiota and fermentation products in equine large intestine in response to dietary variation and intestinal disease. *The British journal of nutrition*, **107**, 989–95.

Daly, K. & Shirazi-Beechey, S.P. (2003) Design and evaluation of group-specific oligonucleotide probes for quantitative analysis of intestinal ecosystems: Their application to assessment of equine colonic microflora. *FEMS Microbiology Ecology*, **44**, 243–252.

Daly, K., Stewart, C.S., Flint, H.J. & Shirazi-Beechey, S.P. (2001) Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes. *FEMS Microbiology Ecology*, **38**, 141–151.

Davies-Morel, M. (2008) *Equine Reproductive Physiology, Breeding and Stud Management*. 2nd edn. CABI, Wallingford.

Davis, G. (1941) Morphology and division in Tetratoxum unifasciculatum Gassovsky. *Transactions of the American Microscopical Society*, **60**, 441–452.

Deda, O., Gika, H.G., Wilson, I. & Theodoridis, G.A. (2015) An overview of fecal sample preparation for global metabolic profiling. *Journal of Pharmaceutical and Biomedical Analysis*, **113**, 137–150.

Demyttenaere, J. & Kimpe, N. De. (2001) Biotransformation of terpenes by fungi study of the pathways involved. *Journal of Molecular Catalysis - B Enzymatic*, **11**, 265–270.

Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., *et al.* (2016) A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell*, **167**, 1339-1353.e21.

Dicks, L.M.T. & Botes, M. (2010) Probiotic lactic acid bacteria in the gastro-intestinal tract: Health benefits, safety and mode of action. *Beneficial Microbes*, **1**, 11–29.

Dicks, L.M.T., Botha, M., Loos, B. & Smith, C. (2015) Adhesion of Lactobacillus reuteri strain Lr1 to equine epithelial cells and competitive exclusion of Clostridium difficile from the gastro-intestinal tract of horses. *Annals of Microbiology*, **65**, 1087–1096.

Dierendonck, M.C. Van, Sigurjónsdóttir, H., Colenbrander, B. & Thorhallsdóttir, a. G. (2004) Differences in social behaviour between late pregnant, post-partum and barren mares in a herd of Icelandic horses. *Applied Animal Behaviour Science*, **89**, 283–297.

Dillon, R.J., Vennard, C.T., Buckling, A. & Charnley, A.K. (2005) Diversity of locust gut bacteria protects against pathogen invasion. *Ecology Letters*, **8**, 1291–1298.

Dixon, E., Clubb, C., Pittman, S., Ammann, L., Rasheed, Z., Kazmi, N., *et al.* (2011) Solid-phase microextraction and human fecal VOC metabolome. *PloS one*, **6**, e18471.

Dougal, K., Blackmore, T., Pachebat, J., Harris, P. & Newbold, C. (2011a) Comparison of bacterial populations from the caecum, right dorsal colon and faeces of horses' using terminal restriction fragment length polymorphism (TRFLP). *Journal of Equine Veterinary Science*, **31**, 261–262.

Dougal, K., Harris, P. a, Edwards, A., Pachebat, J. a, Blackmore, T.M., Worgan, H.J., *et al.* (2012) A comparison of the microbiome and the metabolome of different regions of the equine hindgut. *FEMS Microbiology Ecology*, **82**, 642–52.

Dougal, K., Harris, P.A., Girdwood, S.E., Creevey, C.J., Curtis, G.C., Barfoot, C.F., *et al.* (2017) Changes in the total fecal bacterial population in individual horses aintained on a restricted diet over 6 weeks. *Frontiers in Microbiology*, **8**, doi: 10.3389/fmicb.2017.01502.

Dougal, K., Harris, P.A., Pachebat, J. & Newbold, C. (2011b) Comparison of bacterial populations from different regions of the large intestine of horses using terminal restriction fragment length polymorphism and 454 pyrosequencing. In *Advances in Animal Biosciences*. p. 371.

Dougal, K., La Fuente, G. De, Harris, P. a., Girdwood, S.E., Pinloche, E., Geor, R.J., *et al.* (2014) Characterisation of the faecal bacterial community in adult and elderly horses fed a high fibre, high oil or high starch diet using 454 pyrosequencing. *PLoS ONE*, **9**, e87424.

Dougal, K., la Fuente, G. de, Harris, P. a, Girdwood, S.E., Pinloche, E. & Newbold, C.J. (2013) Identification of a core bacterial community within the large intestine of the horse. *PloS one*, **8**, e77660.

Doxey, B.D.L., Robbt, J. & Milne, E.M. (1990) Mycological studies on the equine intestinal tract with particular reference to equine dysautonomia (grass sickness). *Annals of Applied Biology*, **117**, 337–341.

Drabovich, A., Pavlou, M., Batruch, I. & Diamandis, E. (2013) Proteomic and Mass Spectrometry Technologies for Biomarker Discovery. In *Proteomic and Metabolomic Approaches to Biomarker Discovery* (ed. by Issaq, H. & Veenstra, T.). pp. 17–37.

Drogemuller, M., Beelitz, P., Pfister, K., Schnieder, T. & Samson-Himmelstjerna, G. von. (2004) Amplification of ribosomal DNA of Anoplocephalidae: Anoplocephala perfoliata diagnosis by PCR as a possible alternative to coprological methods. *Veterinary parasitology*, **124**, 205–15.

Duarte, A.M., Jenkins, T.P., Latrofa, M.S., Giannelli, A., Papadopoulos, E., Carvalho, L.M. De, *et al.* (2016) Helminth infections and gut microbiota - A feline perspective. *Parasites and Vectors*, **9**, 1–9.

Duncan, J. & Love, S. (1991) Preliminary observations on an alternative strategy for the control of horse strongyles. *Equine Veterinary Journal*, **23**, 226–228.

Dunn, W.B. & Ellis, D.I. (2005) Metabolomics: Current analytical platforms and methodologies. *Trends in Analytical Chemistry*, **24**, 285–294.

Dunn, W.B., Wilson, I.D., Nicholls, A.W. & Broadhurst, D. (2012) The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*, **4**, 2249–2264.

Dyce, K., Sack, W. & Wensing, C. (2018) The Abdomen of the Horse. In *Textbook of Veterinary Anatomy* (ed. by Singh, B.). Saunders, Philadelphia, pp. 535–551.

Earing, J.E. (2012) Bacterial Colonization of the Equine Gut; Comparison of Mare and Foal Pairs by PCR-DGGE. *Advances in Microbiology*, **02**, 79–86.

Easton, S., Pinchbeck, G.L., Tzelos, T., Bartley, D.J., Hotchkiss, E., Hodgkinson, J.E., *et al.* (2016) Investigating interactions between UK horse owners and prescribers of anthelmintics. *Preventive Veterinary Medicine*, **135**, 17–27.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Edwards, J.E. (2019) Equine anaerobic fungi: Key taxa of central importance to dietary fibre degradation. In *European Equine Health & Nutrition Congress 9th Edition*. pp. 23–31.

Edwards, J.E., Hermes, G.D.A., Kittelmann, S., Nijsse, B. & Smidt, H. (2019) Assessment of the accuracy of high-throughput sequencing of the ITS1 region of Neocallimastigomycota for community composition analysis. *Frontiers in Microbiology*, **10**, doi: 10.3389/fmicb.2019.02370.

Edwards, J.E., Kingston-Smith, A.H., Jimenez, H.R., Huws, S.A., Skøt, K.P., Griffith, G.W., *et al.* (2008) Dynamics of initial colonization of nonconserved perennial ryegrass by anaerobic fungi in the bovine rumen. *FEMS Microbiology Ecology*, **66**, 537–545.

Effmert, U., Kalderás, J., Warnke, R. & Piechulla, B. (2012a) Volatile mediated interactions between bacteria and fungi in the soil. *Journal of Chemical Ecology*, **38**, 665–703.

Effmert, U., Kalderás, J., Warnke, R. & Piechulla, B. (2012b) Volatile Mediated Interactions Between Bacteria and Fungi in the Soil. *Journal of Chemical Ecology*, **38**, 665–703.

Egan, C.E., Snelling, T.J. & Mcewan, N.R. (2010) The onset of ciliate populations in newborn foals. *Acta Protozoologica*, **49**, 145–147.

Elinav, E., Strowig, T., Kau, A.L., Henao-Mejia, J., Thaiss, C.A., Booth, C.J., *et al.* (2011) NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell*, **145**, 745–757.

Ellis, A. & Hill, J. (2005) Digestive physiology of the horse. In *Nutritional Physiology of the Horse* (ed. by Ellis, A. & Hill, J.). Nottingham University Press, Nottingham, pp. 7–43.

Engelhardt, W. Von, Ronnau, K., Rechkemmer, G. & Sakata, T. (1989) Absorption of short-chain fatty acids and their role in the hindgut of mongastric animals. *Animal Feed Science and Technology*, **23**, 43–53.

Engell-Sørensen, K., Pall, A., Damgaard, C. & Holmstrup, M. (2018) Seasonal variation in the prevalence of equine tapeworms using coprological diagnosis during a seven-year period in Denmark. *Veterinary Parasitology: Regional Studies and Reports*, **12**, 22–25.

Ericsson, A.C., Johnson, P.J., Lopes, M.A., Perry, S.C. & Lanter, H.R. (2016) A microbiological map of the healthy equine gastrointestinal tract. *PLoS ONE*, **11**, 1–17.

Escalona, E.E., Leng, J., Dona, a C., Merrifield, C. a, Holmes, E., Proudman, C.J., *et al.* (2014) Dominant components of the Thoroughbred metabolome characterised by (1) H-nuclear magnetic resonance spectroscopy: A metabolite atlas of common biofluids. *Equine veterinary journal*, **47**, 721–730.

Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M., *et al.* (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, **2**, 6.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., *et al.* (2013) The long-term stability of the human gut microbiota. *Science*, **341**, DOI: 10.1126/science.1237439.

Faubladier, C., Sadet-Bourgeteau, S., Philippeau, C., Jacotot, E. & Julliand, V. (2014) Molecular monitoring of the bacterial community structure in foal feces pre- and post-weaning. *Anaerobe*, **25**, 61–6.

Fernandes, K. a, Kittelmann, S., Rogers, C.W., Gee, E.K., Bolwell, C.F., Bermingham, E.N., *et al.* (2014) Faecal microbiota of forage-fed horses in New Zealand and the population dynamics of microbial communities following dietary change. *PloS one*, **9**, e112846.

Ferran, A.A., Bibbal, D., Pellet, T., Laurentie, M., Gicquel-Bruneau, M., Sanders, P., *et al.* (2013) Pharmacokinetic/pharmacodynamic assessment of the effects of parenteral administration of a fluoroquinolone on the intestinal microbiota: Comparison of bactericidal activity at the gut versus the systemic level in a pig model. *International Journal of Antimicrobial Agents*, **42**, 429–435.

Fiehn, O. (2002) Metabolomics - the link between genotyopes and phenotypes. *Plant Molecular Biology*, **48**, 155–171.

Fischer, S., Trefz, P., Bergmann, A., Steffens, M., Ziller, M., Miekisch, W., *et al.* (2015) Physiological variability in volatile organic compounds (VOCs) in exhaled breath and released from faeces due to nutrition and somatic growth in a standardized caprine animal model. *Journal of Breath Research*, **9**, 027108.

Fliegerova, K., Mura, E., Mrázek, J. & Moniello, G. (2016) A comparison of microbial profiles of different regions of the equine hindgut. *Livestock Science*, **190**, 16–19.

Fombelle, A. de, Julliand, V., Drogoul, C. & Jacotot, E. (2001) Feeding and microbial disorders in horses: 1-effects of an abrupt incorporation of two levels of barley in a hay diet on microbial profile and activities. *Journal of Equine Veterinary Science*, **21**, 439–445.

Fombelle, A. De, Varloud, M., Goachet, A., Jacotot, E., Philippeau, C. & Drogoul, C. (2003)

Characterization of the microbial and biochemical profile of the different segments of the digestive tract in horses given two distinct. *Animal Science*, **2**, 293–304.

Frandson, R. (2003) Physiology of Digestion. In *Anatomy and Physiology of Farm Animals* (ed. by Frandson, R., Wilke, W. & Fails, A.). Wiley-Blackwell, Oxford, pp. 329–342.

Frape, D. (2010) *Equine Nutrition and Feeding*. 4th Ed. Wiley-Blackwell, Oxford.

French, P., Stanton, C., Lawless, F., O'Riordan, E., Monahan, F., Caffrey, P., *et al.* (2000) Fatty acid composition , including conjugated linoleic acid , of intramuscular fat from steers offered grazed grass, grass silage, or concentrate-based diets. *Journal of Animal Science*, **78**, 2849–2855.

Gao, X., Cao, Q., Cheng, Y., Zhao, D., Yang, H., Wu, Q., *et al.* (2018) Chronic stress promotes colitis by disturbing the gut microbiota and triggering immune system response. *PNAS*, **115**, E2960–E2969.

Garner, C.E., Smith, S., Elviss, N.C., Humphrey, T.J., White, P., Ratcliffe, N.M., *et al.* (2008) Identification of Campylobacter infection in chickens from volatile faecal emissions. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals*, **13**, 413–21.

Garner, C.E., Smith, S., Lacy Costello, B. de, White, P., Spencer, R., Probert, C.S.J., *et al.* (2007) Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *FASEB Journal*, **21**, 1675–88.

Garner, H., Hutcheson, D., Coffman, J., Hahn, A. & Salem, C. (1977) Lactic acidosis: a factor associated with equine laminitis. *Journal of Animal Science*, **45**, 1037–1041.

Garret, L., Brown, R. & Poxton, I. (2002) A comparative study of the intestinal microbiota of healthy horses and those suffering from equine grass sickness. *Veterinary Microbiology*, **87**, 81–88.

Gasser, R.B., Williamson, R.M.C. & Beveridge, I. (2005) Anoplocephala perfoliata of horses – significant scope for further research, improved diagnosis and control. *Parasitology*, **131**, 1–13.

Gelberg, H. (2017) Chapter 7: Alimentary system and the peritoneum, omentum, mesentery and peritoneal cavity. In *Pathologic Basis of Veterinary Disease* (ed. by Zachary, J.). Elsevier, Missouri, pp. 324–411.

Gergócs, V. & Hufnagel, L. (2009) Application of oribatid mites as indicators. *Applied Ecology and Environmental Research*, **7**, 79–98.

Ghali, M., Scott, P., Alhadrami, G. & Jassim, R. AL. (2011) The predominant lactic acid producing and utilizing bacteria from the gastrointestinal tract of the dromedary camel (Camelus dromedarius) in Australia. *Animal Production Science*, **51**, 587–604.

Gianelli, M.P., Flores, M. & Toldra, F. (2002) Optimisation of solid phase microextraction (SPME) for the analysis of volatile compounds in dry-cured ham. *Journal of the Science of Food and Agriculture*, **82**, 1703–1709.

Giorgio, R. De, Guerrini, S., Barbara, G., Stanghellini, V., Ponti, F. De, Corinaldesi, R., *et al.* (2004) Inflammatory neuropathies of the enteric nervous system. *Gastroenterology*, **126**, 1872–1883.

Girones, R., Ferrus, M., Alonso, J., Rodriguez-Manzano, J., Calgua, B., Abreu Correa, A. de, *et al.* (2010) Molecular detection of pathogens in water - the pros and cons of molecular

techniques. *Water Research*, **44**, 4325–4339.

Göçmen, B., Gürellİ, G. & Dehority, B.A. (2012) Fecal ciliate composition of Cypriot domestic horses (Equus caballus Linnaeus , 1758), **36**, 163–170.

Gold, J., Heath, B. & Bauchop, T. (1988) Ultrastructural description of a new chytrid genus of caecum anaerobe, Caecomyces equi gen. nov., sp. nov., assigned to the Neocallimasticaceae. *BioSystems*, **21**, 403–415.

González, I., Cao, K.A.L., Davis, M.J. & Déjean, S. (2012) Visualising associations between paired "omics" data sets. *BioData Mining*, **5**, 1–23.

Gordon, G. & Phillips, M. (1993) Removal of anaerobic fungi from the rumen of sheep by chemical treatment and the effect on feed consumption and in vivo fibre digestion. *Letters in Applied Microbiology*, **17**, 220–223.

Gorzelak, M.A., Gill, S.K., Tasnim, N., Ahmadi-Vand, Z., Jay, M. & Gibson, D.L. (2015) Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. *PLoS ONE*, **10**, 1–14.

Gratton, J., Phetcharaburanin, J., Mullish, B.H., Williams, H.R.T., Thursz, M., Nicholson, J.K., *et al.* (2016) Optimized sample handling strategy for metabolic profiling of human feces. *Analytical Chemistry*, **88**, 4661–4668.

Greiner, E. & Lane, T. (1994) Effects of the daily feeding of pyrantel tartrate on Anoplocephala infections in three horses: a pilot study. *Journal of Equine Veterinary Science*, **14**, 43–44.

Greiner, M., Pfeiffer, D. & Smith, R.D. (2000) Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*, **45**, 23–41.

Gromski, P.S., Muhamadali, H., Ellis, D.I., Xu, Y., Correa, E., Turner, M.L., *et al.* (2015) A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta*, **879**, 10–23.

Gruninger, R.J., Puniya, A.K., Callaghan, T.M., Edwards, J.E., Youssef, N., Dagar, S.S., *et al.* (2014) Anaerobic fungi (phylum Neocallimastigomycota): Advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiology Ecology*, **90**, 1–17.

Gu, Y.-Q., Mo, M.-H., Zhou, J.-P., Zou, C.-S. & Zhang, K.-Q. (2007) Evaluation and identification of potential organic nematicidal volatiles from soil bacteria. *Soil Biology and Biochemistry*, **39**, 2567–2575.

Gürelli, G. & Göçmen, B. (2011) Intestinal ciliate composition found in the feces of the Turk rahvan horse Equus caballus, Linnaeus 1758. *European journal of protistology*, **47**, 245–55.

Gürelli, G. & Göçmen, B. (2012) Intestinal ciliate composition found in the feces of racing horses from Izmir, Turkey. *European journal of protistology*, **48**, 215–26.

Gylswyk, N. Van & Roche, C. (1970) Characteristics of Ruminococcus and cellulolytic Butyrivibrio species from the rumens of sheep fed differently supplemented Teff (Eragrostis tef) hay diets. *Journal of General Microbiology*, **64**, 11–17.

Handl, S., German, A.J., Holden, S.L., Dowd, S.E., Steiner, J.M., Heilmann, R.M., *et al.* (2013) Faecal microbiota in lean and obese dogs. *FEMS microbiology ecology*, **84**, 332–43.

Hansen, N., Avershina, E., Mydland, L., Naesset, J., Austbø, D., Moen, B., *et al.* (2015) High nutrient availability reduces the diversity and stability of the equine caecal microbiota. *Microbial Ecology in Health and Disease*, **1**, 1–8.

Hastie, P., Mitchell, K. & Murray, J. (2008) Semi-quantitative analysis of Ruminococcus flavefaciens, Fibrobacter succinogens and Streptococcus bovis in the equine large intestine using real-time polymerase chain reaction. *British Journal of Nutrition*, **107**, 916–995.

Hayes, K., Bancroft, A., Goldrick, M., Portsmouth, C., Roberts, I. & Grencis, R. (2010) Exploitation of the Intestinal Microflora by the Parasitic Nematode Trichuris muris. *Science*, **328**, 1391–139.

Head, I.M., Saunders, J.R. & Pickup, R.W. (1998) Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. *Microbial Ecology*, **35**, 1–21.

Hector, M. & Linden, R. (1999) Reflexes of salivary secretion. In *Neural Mechanisms of Salivary Secretion* (ed. by Garret, J., Elkstrom, J. & Anderson, L.). Karger, Basel, pp. 196–217.

Heidmann, P., Saulez, M. & Cebra, C. (2004) Pyloric stenosis with reflux oesophagitis in a Thoroughbred filly. *Equine Veterinary Education*, **16**, 172–177.

Hillyer, M., Taylor, F. & French, N. (2001) A cross-sectional study of colic in horses on thoroughbred training premises in the British Isles in 1997. *Equine Veterinary Journal*, **33**, 380–385.

Hillyer, M.H., R Taylor, F.G., Proudman, C.J., Edwards, G.B., Smith, J.E. & French, N.P. (2002) Case control study to identify risk factors for simple colonic obstruction and distension colic in horses. *Equine Veterinary Journal*, **34**, 455–463.

Hinton, A. & Hume, M. (1995) Antibacterial activity of the metabolic by-products of a Veilonella species and Bacteroides fragilis. *Anaerobe*, **1**, 121–127.

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G.D., *et al.* (2013) Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS ONE*, **8**.

Hoglund, J., Ljungstrm, B., Nilsson, O. & Uggla, A. (1995) Enzyme-linked immunosorbent assay (ELISA) for the detection of antibodies to Anoplocephala perfoliata in horse sera. *Veterinary Parasitology*, **59**, 97–106.

Holcombe, S.J., Embertson, R.M., Kurtz, K. a, Roessner, H. a, Wismer, S.E., Geor, R.J., *et al.* (2014) Increased serum nonesterified fatty acid and low ionised calcium concentrations are associated with post partum colic in mares. *Equine Veterinary Journal*, **48**, 39–44.

Hook, S., Wright, A. & McBride, B. (2010) Methanogens: methane producers of the rumen and mitigation strategies. *Archaea*, **2010**, doi:10.1155/2010/945785.

Hopkins, M.J. & Macfarlane, G.T. (2002) Changes in predominant bacterial populations in human faeces with age and with Clostridium difficile infection. *Journal of Medical Microbiology*, **51**, 448–454.

Houlden, A., Hayes, K.S., Bancroft, A.J., Worthington, J.J., Wang, P., Grencis, R.K., *et al.* (2015) Chronic Trichuris muris infection in C57BL/6 mice causes significant changes in host microbiota and metabolome: Effects reversed by pathogen clearance. *PLoS ONE*, **10**, 10.1371/journal.pone.0125945.

Houpt, K. (2002) Formation and dissolution of the mare-foal bond. *Applied Animal*

Behaviour Science, **78**, 319–328.

Hreinsdóttir, I., Hreinsdóttir, A., Eydal, M., Tysnes, K.R. & Robertson, L.J. (2019) Anoplocephala perfoliata infection in horses in iceland: investigation of associations between intensity of infection and lesions. *Journal of Parasitology*, **105**, 379.

Hsiung, T. (1930) Some new ciliates from the large intestine of the horse. *Transactions of the American Microscopical Society*, **39**, 34–41.

Hughes, N.C., Wong, E.Y.K., Fan, J. & Bajaj, N. (2007) Determination of carryover and contamination for mass spectrometry-based chromatographic assays. *The AAPS journal*, **9**, E353-60.

Hunter, L., Miller, J. & Poxton, I. (1999) The association of Clostridium botulinum type C with equine grass sickness: a toxicoinfection? *Equine veterinary Journal*, **31**, 492–499.

Hwang, O.H., Cho, S.B., Han, D.W., Lee, S.R., Kwag, J.H. & Park, S.K. (2016) Effect of Storage Period on the Changes of Odorous Compound Concentrations and Bacterial Ecology for Identifying the Cause of Odor Production from Pig Slurry. *Plos One*, **11**, e0162714.

Ishizaka, S., Matsuda, A., Amagai, Y., Oida, K., Jang, H., Ueda, Y., *et al.* (2014) Oral Administration of Fermented Probiotics Improves the Condition of Feces in Adult Horses. *Journal of Equine Science*, **25**, 65–72.

Jalanka-Tuovinen, J., Salonen, A., Nikkilä, J., Immonen, O., Kekkonen, R., Lahti, L., *et al.* (2011) Intestinal microbiota in healthy adults: Temporal analysis reveals individual and common core and relation to intestinal symptoms. *PLoS ONE*, **6**, 10.1371/journal.pone.0023035.

Jassim, R. a M. Al & Andrews, F.M. (2009) The bacterial community of the horse gastrointestinal tract and its relation to fermentative acidosis, laminitis, colic, and stomach ulcers. *The Veterinary clinics of North America. Equine practice*, **25**, 199–215.

Jassim, R.A.M. Al, Scott, P.T., Trebbin, A.L., Trott, D. & Pollitt, C.C. (2005) The genetic diversity of lactic acid producing bacteria in the equine gastrointestinal tract. *FEMS Microbiology Letters*, **248**, 75–81.

Jawhara, S., Thuru, X., Standaert-Vitse, A., Jouault, T., Mordon, S., Sendid, B., *et al.* (2008) Colonization of mice by candida albicans is promoted by chemically induced colitis and augments inflammatory responses through Galectin-3. *The Journal of Infectious Diseases*, **197**, 972–980.

Jenkins, T.P., Peachey, L.E., Ajami, N.J., MacDonald, A.S., Hsieh, M.H., Brindley, P.J., *et al.* (2018) Schistosoma mansoni infection is associated with quantitative and qualitative modifications of the mammalian intestinal microbiota. *Scientific Reports*, **8**, 12072.

Jochmann, M., Laaks, J. & Schmidt, T. (2014) Solvent-free extraction and injection techniques. In *Practical Gas Chromatography* (ed. by Dettmer-Wilde, K. & Engewald, W.). Springer, London, pp. 372–393.

Johnson, W.E. & Li, C. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.

Joshi, N. & Fass, J. (2011) A sliding-window, adaptive, quality-based trimming tool for FastQ files [WWW Document]. *https://github.com/najoshi/sickle*. URL [accessed on 2011].

Jost, T., Lacroix, C., Braegger, C. & Chassard, C. (2014) Stability of the maternal gut microbiota during late pregnancy and early lactation. *Current Microbiology*, **68**, 419–427.

Julliand, V. (2005) Impact of nutrition on the microflora of the gastro-intestinal tract in horses. In *Applied Equine Nutrition: Equine Nutrition Conference (ENUCO)* (ed. by Lindner, A.). Wageningen Academic Publishers, The Netherlands, p. 85.

Julliand, V., Riondet, C., Vaux, A. de, Alcaraz, G. & Fonty, G. (1998) Comparison of metabolic activities between Piromyces citronii, an equine fungal species, and Piromyces communis, a ruminal species. *Animal Feed Science and Technology*, **70**, 161–168.

Julliand, V., Vaux, A. De, Villard, L. & Richard, Y. (1996) Preliminary studies on the bacterial flora of faeces taken from foals, from birth to twelve weeks. Effect of the oral administration of a commercial colostrum replacer. *Pferdeheikunde*, **12**, 209–212.

Julliand, V., Vaux, A., Millet, L. & Fonty, G. (1999) Identification of Ruminococcus flavefaciens as the predominant cellulolytic bacterial species of the equine caecum. *Applied Environmental Microbiology*, **65**, 3738–3741.

Kai, M., Haustein, M., Molina, F., Petri, A., Scholz, B. & Piechulla, B. (2009) Bacterial volatiles and their action potential. *Applied Microbiology and Biotechnology*, **81**, 1001–1012.

Kaplan, R.M. & Nielsen, M.K. (2010) An evidence-based approach to equine parasite control: It ain't the 60s anymore. *Equine Veterinary Education*, **22**, 306–316.

Karu, N., Deng, L., Slae, M., Guo, C., Sajed, T., Huynh, H., *et al.* (2018) A review on human fecal metabolomics: methods, applications and the human fecal metabolome database. *Analytica Chimica Acta*, **1030**, 1–24.

Kennedy, M. (2011) Latherin and other biocompatible surfactant proteins. *Biochemical Society Transactions*, **39**, 1017–1022.

Kern, D., Slyter, L., Leffel, E., Weaver, J. & Oltjen, R. (1974) Ponies vs. steers: microbial and chemical characteristics of intestinal ingesta. *Journal of Animal Science*, **38**, 559–564.

Khalid, T., Aggio, R., White, P., Lacy Costello, B. De, Persad, R., Al-Kateb, H., *et al.* (2015) Urinary volatile organic compounds for the detection of prostate cancer. *PLoS ONE*, **10**, 1–15.

Khan, W.I., Blennerhasset, P.A., Varghese, A.K., Chowdhury, S.K., Omsted, P., Deng, Y., *et al.* (2002) Intestinal nematode infection ameliorates experimental colitis in mice. *Infection and Immunity*, **70**, 5931–5937.

Kieler, I.N., Mølbak, L., Hansen, L.L., Hermann-Bank, M.L. & Bjornvad, C.R. (2016) Overweight and the feline gut microbiome - a pilot study. *Journal of Animal Physiology and Animal Nutrition*, **100**, 478–484.

Kjaer, L.N., Lungholt, M.M., Nielsen, M.K., Olsen, S.N. & Maddox-Hyttel, C. (2007) Interpretation of serum antibody response to Anoplocephala perfoliata in relation to parasite burden and faecal egg count. *Equine veterinary journal*, **39**, 529–33.

Kobayashi, Y., Koike, S., Miyaji, M., Hata, H. & Tanaka, K. (2006) Hindgut microbes, fermentation and their seasonal variations in Hokkaido native horses compared to light horses. *Ecological Research*, **21**, 285–291.

Koike, S., Shingu, Y., Inaba, H., Kawai, M., Kobayashi, Y., Hata, H., *et al.* (2000) Fecal Bacteria in Hokkaido Native Horses as Characterized by Microscopic Enumeration and Competitive Polymerase Chain Reaction Assays. *Journal of Equine Science*, **11**, 45–50.

Koning, S., Janssen, H.-G. & Brinkman, U.A.T. (2009) Modern methods of sample preparation for GC analysis. *Chromatographia*, **69**, 33–78.

Koren, O., Goodrich, J.K., Cullender, T.C., Spor, A., Laitinen, K., Bäckhed, H.K., *et al.* (2012) Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell*, **150**, 470–80.

Korpi, A., Järnberg, J. & Pasanen, A.L. (2009) Microbial volatile organic compounds. *Critical Reviews in Toxicology*, **39**, 139–193.

Kovacs, A., Ben-Jacob, N., Tayem, H., Halperin, E., Iraqi, F. & Gophna, U. (2011) Genotype is a stronger determination than sex of the mouse gut microbiota. *Environmental Microbiology*, **61**, 423–428.

Krause, D.O., Nagaraja, T.G., Wright, A.D.G. & Callaway, T.R. (2013) Board-invited review: Rumen microbiology: Leading the way in microbial ecology. *Journal of Animal Science*, **91**, 331–341.

Krehbiel, C.R. (2014) Invited Review: Applied nutrition of ruminants: Fermentation and digestive physiology. *Professional Animal Scientist*, **30**, 129–139.

Kreisinger, J., Bastien, G., Hauffe, H.C., Marchesi, J. & Perkins, S.E. (2015) Interactions between multiple helminths and the gut microbiota in wild rodents. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **370**, 20140295.

Kristoffersen, C., Jensen, R.B., Avershina, E., Austbø, D., Tauson, A.-H. & Rudi, K. (2016) Diet-dependent modular dynamic interactions of the equine cecal microbiota. *Microbes and Environments*, **31**, 378–386.

Kruk, J., Doskocz, M., Jodlowska, E., Zacharzewska, A., Lakomiec, J., Czaja, K., *et al.* (2016) NMR Techniques in Metabolomic Studies: A Quick Overview on Examples of Utilization. *Applied Magnetic Resonance*, **48**, 1–21.

Kumar, S., Indugu, N., Vecchiarelli, B. & Pitta, D.W. (2015) Associative patterns among anaerobic fungi, methanogenic archaea, and bacterial communities in response to changes in diet and age in the rumen of dairy cows. *Frontiers in Microbiology*, **6**, 1–10.

la Fuente, G. de, Belanche, A., Girwood, S.E., Pinloche, E., Wilkinson, T. & Newbold, C.J. (2014) Pros and cons of ion-torrent next generation sequencing versus terminal restriction fragment length polymorphism T-RFLP for studying the rumen bacterial community. *PloS one*, **9**, e101435.

Lacy Costello, B. de, Amann, a, Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., *et al.* (2014) A review of the volatiles from the healthy human body. *Journal of breath research*, **8**, 014001.

Laho, T., Varadyova, Z., Mihalikova, K. & Kisidayova, S. (2013) Fermentation Capacity of Fecal Microbial Inocula of Przewalski Horse, Kulan, and Chapman Zebra and Polysaccharide Hydrolytic Activities of Fecal Microbial Constituents (Ciliates and Bacteria) of Kulan and Chapman Zebra. *Journal of Equine Veterinary Science*, **33**, 143–149.

Laor, Y., Shabtay, A., Ravid, U., Baybikov, R. & Eitam, H. (2007) Changes in VOCs Emissions from Fecal Manure throughout the Life Cycle of Beef Cattle. In *ASABE Annual International Meeting*. p. 074003.

Larsen, M., Nansen, P., Henriksen, S.A., Wolstrup, J., Grønvold, J., Zorn, A., *et al.* (1995) Predacious activity of the nematode-trapping fungus Duddingtonia flagrans against cyathostome larvae in faeces after passage through the gastrointestinal tract of horses. *Veterinary Parasitology*, **60**, 315–320.

Law, R. & Morton, R. (1996) Permanence and the assembly of ecological communities.

Ecology, **77**, 762–775.

Lê Cao, K.A., Boitard, S. & Besse, P. (2011) Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*, **12**, 253.

Lee, D.L. & Tatchell, R.J. (1964) Studies on the Tapeworm Anoplocephala Perfoliata (Goeze, 1782). *Parasitology*, **54**, 467–79.

Lee, S.C., Tang, M.S., Lim, Y. a L., Choy, S.H., Kurtz, Z.D., Cox, L.M., *et al.* (2014) Helminth colonization is associated with increased diversity of the gut microbiota. *PLoS neglected tropical diseases*, **8**, e2880.

Leng, J., Escalona, E., Blow, F., Proudman, C. & Swann, J. (2014) Comparison of the faecal bacteria population of equine grass sickness and matched controls by next generation sequencing of bacterial DNA. In *Abstract Programme: 11th International Equine Colic Research Symposium*. Dublin, p. https://docplayer.net/13449869-Abstract-programme-.

Leng, J., Proudman., C., Blow, F., Darby, A. & Swann, J. (2015) Undertstanding intestinal microbiota in equine grass sickness: next generation sequencing of faecal bacterial DNA. *Equine Veterinary Journal, Speciial Issue: Clinical Research Abstracts of the British Equine Veterinary Association Congress 2015*, **47**, 9.

Leser, T.D. & Mølbak, L. (2009) Better living through microbial action: the benefits of the mammalian gastrointestinal microbiota on the host. *Environmental microbiology*, **11**, 2194–206.

Lester, H., Morgan, E., Hodgkinson, J. & Matthews, J. (2018) Analysis of strongyle egg sheding consistency in horses and factors that affect it. *Journal of Equine Veterinary Science*, **60**, 113–119.

Li, J. V, Saric, J., Wang, Y., Keiser, J., Utzinger, J. & Holmes, E. (2011a) Chemometric analysis of biofluids from mice experimentally infected with Schistosoma mansoni. *Parasites & Vectors*, **4**, 179.

Li, R.W., Li, W., Sun, J., Yu, P., Baldwin, R.L. & Urban, J.F. (2016) The effect of helminth infection on the microbial composition and structure of the caprine abomasal microbiome. *Scientific Reports*, **6**, 1–10.

Li, R.W., Wu, S., Li, W., Huang, Y. & Gasbarre, L.C. (2011b) Metagenome plasticity of the bovine abomasal microbiota in immune animals in response to Ostertagia ostertagi infection. *PloS one*, **6**, e24417.

Li, R.W., Wu, S., Li, W., Navarro, K., Couch, R.D., Hill, D., *et al.* (2012) Alterations in the porcine colon microbiota induced by the gastrointestinal nematode Trichuris suis. *Infection and Immunity*, **80**, 2150–7.

Liggenstoffer, A.S., Youssef, N.H., Couger, M.B. & Elshahed, M.S. (2010) Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores. *ISME Journal*, **4**, 1225–1235.

Lightbody, K.L., Davis, P.J. & Austin, C.J. (2016) Validation of a novel saliva-based ELISA test for diagnosing tapeworm burden in horses. *Veterinary Clinical Pathology*, **45**, 335–346.

Lightbody, K.L., Matthews, J.B., Kemp-Symonds, J.G., Lambert, P.A. & Austin, C.J. (2018) Use of a saliva-based diagnostic test to identify tapeworm infection in horses in the UK. *Equine Veterinary Journal*, **50**, 213–219.

Ligtenberg, A.J.M., Brand, H.S., Keijbus, P.A.M. Van Den & Veerman, E.C.I. (2015) The effect of physical exercise on salivary secretion of MUC5B, amylase and lysozyme. *Archives of Oral Biology*, **60**, 1639–1644.

Lin, C. & Stahl, D.A. (1995) Taxon-specific probes for the cellulolytic genus Fibrobacter reveal abundant and novel equine-associated populations. *Applied and Environmental Microbiology*, **61**, 1348–1351.

Lin, J., Aoll, J., Niclass, Y., Ine, M., Wu, L., Pika, J., *et al.* (2013) Qualitative and Quantitative Analysis of Volatile Constituents from Latrines. *Environmental Science Technology*, **47**, 7876–7882.

Little, D. & Blikslager, A.T. (2002) Factors associated with development of ileal impaction in horses with surgical colic: 78 cases (1986-2000). *Equine Veterinary Journal*, **34**, 464–468.

Liu, C., Kachur, S., Dwan, M., Abraham, A., Aziz, M., Hsueh, P.-R., *et al.* (2012) FungiQuant: A broad-coverage fungal quantitative real-time PCR assay. *BMC Microbiology*, **12**, 255.

Liu, X., Fan, H., Ding, X., Hong, Z., Nei, Y., Liu, Z., *et al.* (2014) Analysis of the gut microbiota by high-throughput sequencing of the v5-v6 regions of the 16s rRNA gene in donkey. *Current Microbiology*, **68**, 657–662.

Long, J. & Orlando, R. (1999) Esophageal submucosal glands: structure and function. *American Journal of Gastroenterology*, **41**, 625–630.

Longland, A. & Byrd, B. (2006) Pasture nonstructral carbohydrates and equine laminitis. *The Journal of Nutrition*, **136**, 2099S-2102S.

Louis, P. & Flint, H.J. (2009) Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiology Letters*, **294**, 1–8.

Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 1–21.

Lowder, M. & Mueller, P. (1998) Dental disease in geriatric horses. *Veterinary Clinics of North America: Equine Practice*, **14**, 365–380.

Lwin, K.O. & Matsui, H. (2014) Comparative analysis of the methanogen diversity in horse and pony by using mcra gene and archaeal 16S rRNA Gene clone libraries. *Archaea*, **2014**, 1–11.

Lyons, E., Drudge, J., Tolliver, S., Swerczek, T. & Collins, S. (1989) Determination of the efficacy of pyrantel pamoate at the therapeutic dose rate against the tapeworm Anoplocephala perfoliata in equids using s modification of the critical test method. *Veterinary Parasitology*, **31**, 13–18.

Lyons, E.T., Bellaw, J.L., Dorton, A.R. & Tolliver, S.C. (2017) Efficacy of moxidectin and an ivermectin-praziquantel combination against ascarids, strongyles, and tapeworms in Thoroughbred yearlings in field tests on a farm in Central Kentucky in 2016. *Veterinary Parasitology: Regional Studies and Reports*, **8**, 123–126.

Lyons, E.T., Bolin, D.C., Bryant, U.K., Cassone, L.M., Jackson, C.B., Janes, J.G., *et al.* (2018) Postmortem examination (2016–2017) of weanling and older horses for the presence of select species of endoparasites: Gasterophilus spp., Anoplocephala spp. and Strongylus spp. in specific anatomical sites. *Veterinary Parasitology: Regional Studies and Reports*, **13**, 98–104.

Lyons, E.T., Tolliver, S.C., Stamper, S., Drudge, J.H., Granstrom, D.E. & Collins, S.S. (1995)

Activity of praziquantel (0.5 mg kg-1) against Anoplocephala perfoliata (Cestoda) in equids. *Veterinary Parasitology*, **56**, 255–257.

Maccaferri, S., Biagi, E. & Brigidi, P. (2011) Metagenomics: key to human gut microbiota. *Digestive Diseases*, **29**, 525–530.

Mach, N., Foury, A., Kittelmann, S., Reigner, F., Moroldo, M., Ballester, M., *et al.* (2017) The effects of weaning methods on gut microbiota composition and horse physiology. *Frontiers in Physiology*, **8**, 1–21.

Mach, N. & Fuster-Botella, D. (2017) Endurance exercise and gut microbiota: A review. *Journal of Sport and Health Science*, **6**, 179–197.

Mackie, R. & Wilkins, C. (1988) Enumeration of Anaerobic Bacterial Microflora of the Equine Gastrointestinal Tract. *Applied and Environmental Microbiology*, **54**, 2155–2160.

Maczulak, A.E., Dawson, K.A. & Baker, J.P. (1985) Nitrogen Utilization in Bacterial Isolates from the Equine Cecumt, **50**, 1439–1443.

Mahé, F., Rognes, T., Quince, C., Vargas, C. de & Dunthorn, M. (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, **3**, e1420.

Maizels, R., Pearce, E., Artis, D., Yazdanbakhsh, M. & Wynn, T. (2009) Regulation of pathogenesis and immunity in helminth infections. *The Journal of Experimental Medicine*, **206**, 2059–2066.

Mani, V. (1999) Properties of commercial SPME fibre coatings. In *Applications of Solid Phase Microextraction* (ed. by Pawliszyn, J.). The Royal Society of Chemistry, Cambridge, UK., pp. 57–108.

Marshall, K., Thompson, K.A., Walsh, D.M. & Baxter, G.D. (1998) Incidence of urinary incontinence and constipation during pregnancy and postpartum: survey of current findings at the Rotunda Lying-In Hospital. *British Journal of Obstetrics and Gynaecology*, **105**, 400–402.

Martin, M. (2011) Cutadapt removes adapter sequence from high-throughput sequencing reads. *EBMnet.journal*, **17**, 10–12.

Massacci, F.R., Clark, A., Ruet, A., Lansade, L., Costa, M. & Mach, N. (2020) Inter-breed diversity and temporal dynamics of the faecal microbiota in healthy horses. *Journal of Animal Breeding and Genetics*, **137**, 103–120.

Matthews, J.B., Hodgkinson, J., Dowdall, S. & Proudman, C. (2003) Recent developments in research into the Cyathostominae and Anoplocephala perfoliata. *Veterinary Research*, **35**, 371–381.

Maurice, C.F., Cl Knowles, S., Ladau, J., Pollard, K.S., Fenton, A., Pedersen, A.B., *et al.* (2015) Marked seasonal variation in the wild mouse gut microbiota. *The ISME Journal*, **9**, 1–12.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal*, **6**, 610–618.

McDonald, P., Edwards, R., Greenhalgh, J., Morgan, C., Sinclair, L. & Wilkinson, R. (2011) *Animal Nutrition*. 7th edn. Pearson Education, Essex.

McDonald, R., Fleming, R., Beeley, J., Bovell, D., Lu, J., Zhao, X., *et al.* (2009) Latherin: a surfactant prtein of horse sweat and saliva. *PloS one*, **4**, e5726.

McKenna, P., Hoffmann, C., Minkah, N., Aye, P.P., Lackner, A., Liu, Z., *et al.* (2008) The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS Pathogens*, **4**, e20.

Meana, A., Luzon, M., Corchero, J. & Gmez-bautista, M. (1998) Reliability of coprological diagnosis of Anoplocephala perfoliata infection. *Veterinary Parasitology*, **74**, 79–83.

Meana, A., Pato, N.F., Martin, R., Mateos, A., Perez-Garcia, J. & Luzon, M. (2005) Epidemiological studies on equine cestodes in central Spain: Infection pattern and population dynamics. *Veterinary Parasitology*, **130**, 233–240.

Medina, B., Girard, I.D., Jacotot, E. & Julliand, V. (2002) Effect of a preparation of Saccharomyces cerevisiae on microbial profiles and fermentation patterns in the large intestine of horses fed a high fiber or a high starch diet. *Journal of Animal Science*, **80**, 2600–2609.

Meisel, J.S. & Grice, E.A. (2016) The human microbiome. In *Genomic and Precision Medicine: Foundations, Translation, and Implementation: Third Edition* (ed. by Ginsburg, G. & Willard, H.). Elsevier Inc., pp. 63–77.

Melo, M. De, Araujo, A.C. V., Chogi, M.A.N. & Duarte, I.C.S. (2018) Cellulolytic and lipolytic fungi isolated from soil and leaf litter samples from the Cerrado (Brazilian Savanna). *Revista de Biología Tropical*, **66**, 237.

Merrit, A. (1999) Normal equine gastroduodenal secretion and motility. *Veterinary Journal*, **29**, 7–13.

Milinovich, G.J., Burrell, P.C., Pollitt, C.C., Bouvet, A. & Trott, D.J. (2008) Streptococcus henryi sp. nov. and Streptococcus caballi sp. nov., isolated from the hindgut of horses with oligofructose-induced laminitis. *International Journal of Systematic and Evolutionary Microbiology*, **58**, 262–266.

Milinovich, G.J., Trott, D.J., Burrell, P.C., Croser, E.L., Jassim, R. a M. Al, Morton, J.M., *et al.* (2007) Fluorescence in situ hybridization analysis of hindgut bacteria associated with the development of equine laminitis. *Environmental microbiology*, **9**, 2090–100.

Miller, T.L. & Wolin, M.J. (1996) Pathways of acetate , propionate , and butyrate formation by the human fecal microbial flora. *Applied and Environmental Microbiology*, **62**, 1589–1592.

Mills, G. a & Walker, V. (2000) Headspace solid-phase microextraction procedures for gas chromatographic analysis of biological fluids and materials. *Journal of Chromatography A*, **902**, 267–287.

Miyaji, M., Ueda, K., Kobayashi, Y., Hata, H. & Kondo, S. (2008a) Fiber digestion in various segments of the hindgut of horses fed grass hay or silage. *Animal Science Journal*, **79**, 339–346.

Miyaji, M., Ueda, K., Nakatsuji, H., Tomioka, T., Kobayashi, Y., Hata, H., *et al.* (2008b) Mean retention time of digesta in the different segments of the equine hindgut. *Animal Science Journal*, **79**, 89–96.

Mochalski, P., Wzorek, B., Śliwka, I. & Amann, A. (2009) Suitability of different polymer bags for storage of volatile sulphur compounds relevant to breath analysis. *Journal of Chromatography B*, **877**, 189–196.

Montes, A. & Pugh, D. (1993) The use of probiotics in food-animal practice. *Journal of*

Veterinary Medicine, **45**, 282–288.

Moore, B. & Dehority, B. (1993) Effects of diet and hindgut defaunation on diet digestibility and microbial concentrations in the cecum and colon of the horse. *Journal of Animal Science*, **71**, 3350–3358.

Moraes, C.M. De, Wanjiku, C., Stanczyk, N.M., Pulido, H., Sims, J.W., Betz, H.S., *et al.* (2018) Volatile biomarkers of symptomatic and asymptomatic malaria infection in humans. *Proceedings of the National Academy of Sciences*, **115**, 5780–5785.

Moreau, M.M., Eades, S.C., Reinemeyer, C.R., Fugaro, M.N. & Onishi, J.C. (2014) Illumina sequencing of the V4 hypervariable region 16S rRNA gene reveals extensive changes in bacterial communities in the cecum following carbohydrate oral infusion and development of early-stage acute laminitis in the horse. *Veterinary microbiology*, **168**, 436–41.

Morgan, M.E. & Pereira, R.L. (1962) Volatile Constituents of Grass and Corn Silage. II. Gas-entrained Aroma. *Journal of Dairy Science*, **45**, 467–471.

Morita, H., Shiratori, C., Murakami, M., Takami, H., Kato, Y., Endo, A., *et al.* (2007) Lactobacillus hayakitensis sp. nov., isolated from intestines of healthy thoroughbreds. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 2836–2839.

Morley, J. (2007) The aging gut: physiology. *Gastroenterology*, **23**, 757–767.

Morrison, P.K., Newbold, C.J., Jones, E., Worgan, H.J., Grove-White, D.H., Dugdale, A.H., *et al.* (2018) The equine gastrointestinal microbiome: Impacts of age and obesity. *Frontiers in Microbiology*, **9**, 1–13.

Mshelia, E.S., Adamu, L., Wakil, Y., Turaki, U.A., Gulani, I.A. & Musa, J. (2018) The association between gut microbiome, sex, age and body condition scores of horses in Maiduguri and its environs. *Microbial Pathogenesis*, **118**, 81–86.

Muhonen, S., Julliand, V., Lindberg, J.E., Bertilsson, J. & Jansson, a. (2009) Effects on the equine colon ecosystem of grass silage and haylage diets after an abrupt change from hay. *Journal of animal science*, **87**, 2291–8.

Mura, E., Edwards, J., Kittelmann, S., Kaerger, K., Voigt, K., Mrázek, J., *et al.* (2019) Anaerobic fungal communities differ along the horse digestive tract. *Fungal Biology*, **123**, 240–246.

Murayama, C., Kimura, Y. & Setou, M. (2009) Imaging mass spectrometry: Principle and application. *Biophysical Reviews*, **1**, 131–139.

Murphy, D. & Love, S. (1997) The pathogenic effects of experimental cyathostome infections in ponies. *Veterinary Parasitology*, **70**, 99–110.

Muyzer, G., Brinkoff, T., Nubel, U., Santegoeds, C., Schafer, H. & C, W. (2004) Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. In *Molecular Microbial Ecology Manual* (ed. by Kowalchuk, G., Bruijin, F. de, Head, I., Akkermans, A. & Elsas, J. van). Dordrecht: Kluwer Academic Publishers, pp. 745–770.

Nagpal, D., Prakash, S., Bhat, K. & Singh, G. (2016) Detection and comparison of Selenomonas sputigena in subgingival biofilms in chronic and aggresive periodontitis patients. *Indian Society of Periodontology*, **20**, 286–291.

Naidu, A.S., Bidlack, W.R. & Clemens, R.A. (1999) Probiotic spectra of lactic acid bacteria (LAB). *Critical reviews in food science and nutrition*, **39**, 13–126.

Nielsen, M. (2016) Equine tapeworm infections: Disease, diagnosis and control. *Equine Veterinary Education*, **28**, 388–395.

Nielsen, M., Mittel, L., Grice, A., Erskine, M., Graves, E., Vaala, W., *et al.* (2013) AAEP Parasite Control Guidelines [WWW Document]. *American Association of Equine Practioners*. URL http://www.aaep.org [accessed on 2013].

Nielsen, M.K., Baptiste, K.E., Tolliver, S.C., Collins, S.S. & Lyons, E.T. (2010a) Analysis of multiyear studies in horses in Kentucky to ascertain whether counts of eggs and larvae per gram of feces are reliable indicators of numbers of strongyles and ascarids present. *Veterinary Parasitology*, **174**, 77–84.

Nielsen, M.K., Baptiste, K.E., Tolliver, S.C., Collins, S.S. & Lyons, E.T. (2010b) Analysis of multiyear studies in horses in Kentucky to ascertain whether counts of eggs and larvae per gram of feces are reliable indicators of numbers of strongyles and ascarids present. *Veterinary Parasitology*, **174**, 77–84.

Nielsen, M.K., Reinemeyer, C.R., Donecker, J.M., Leathwick, D.M., Marchiondo, a a & Kaplan, R.M. (2014) Anthelmintic resistance in equine parasites--current evidence and knowledge gaps. *Veterinary parasitology*, **204**, 55–63.

Nilsson, O., Ljungstrom, B.L., Höglund, J., Lundquist, H. & Uggla, A. (1995) Anoplocephala perfoliata in horses in Sweden: prevalence, infection levels and intestinal lesions. *Acta Veterinary Scandanavia*, **36**, 319–328.

Nilsson, R.H., Larsson, K.H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., *et al.* (2019) The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research*, **47**, D259–D264.

Norton, R. & Behan-Pelletier, V. (2009) Suborder Oribatida. In *A Manual to Acarology* (ed. by Krantz, G. & Walter, D.). Texas Tech University Press, Texas., pp. 430–564.

National Research Council. (2007) Proteins and Amino Acids. In *Nutrient Requirements of Horses*. National Academy Press, Washington DC., pp. 235–267.

O'Donnell, M.M., Harris, H.M.B., Jeffery, I.B., Claesson, M.J., Younge, B., O'Toole, P.W., *et al.* (2013) The core faecal bacterial microbiome of Irish Thoroughbred racehorses. *Letters in Applied Microbiology*, **57**, 492–501.

Oniciuc, E.A., Likotrafiti, E., Alvarez-molina, A. & Prieto, M. (2018) Resistant microorganisms and antimicrobial resistance genes across the food chain. *Genes*, **9**, doi:10.3390/genes9050268.

Orcutt, B., Bailey, B., Staudigel, H., Tebo, B.M. & Edwards, K.J. (2009) An interlaboratory comparison of 16S rRNA gene-based terminal restriction fragment length polymorphism and sequencing methods for assessing microbial diversity of seafloor basalts. *Environmental Microbiology*, **11**, 1728–1735.

Orkin, J.D., Campos, F.A., Myers, M.S., Cheves Hernandez, S.E., Guadamuz, A. & Melin, A.D. (2018) Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest. *The ISME Journal*, **13**, 183–196.

Orpin, C.G. (1981) Isolation of cellulolytic phycomycete fungi from the caecum of the horse. *Journal of General Microbiology*, **123**, 287–96.

Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G.A., Papanikolaou, N., Kotoulas, G., *et al.* (2015) Metagenomics: Tools and insights for analyzing next-generation sequencing

data derived from biodiversity studies. *Bioinformatics and Biology Insights*, **9**, 75–88.

Owen, J., McCullagh, K. & Crook, D. (1978) Seasonal variations in the nutrition of horses at grass. *Equine Veterinary Journal*, **10**, 260–266.

Owen, R., Jagger, D. & Quantaylor, R. (1989) Caecal intussusceptions in horses and the significance of Anoplocephala perfoliata. *Veterinary Record*, **124**, 34–37.

Paliy, O. & Agans, R. (2012) Application of phylogenetic microarrays to interrogation of human microbiota. *FEMS Microbiology Ecology*, **79**, 2–11.

Parfrey, L.W., Jirků, M., Šíma, R., Jalovecká, M., Sak, B., Grigore, K., *et al.* (2017) A benign helminth alters the host immune system and the gut microbiota in a rat model system. *PLoS ONE*, **12**, 1–22.

Pavone, S., Veronesi, F., Genchi, C., Fioretti, D.P., Brianti, E. & Mandara, M.T. (2011) Pathological changes caused by Anoplocephala perfoliata in the mucosa/submucosa and in the enteric nervous system of equine ileocecal junction. *Veterinary Parasitology*, **176**, 43–52.

Pawliszyn, J. (2000) Theory of Solid-Phase Microextraction. *Journal of Chromatographic Science*, **38**, 270–278.

Peachey, L.E., Castro, C., Molena, R.A., Jenkins, T.P., Griffin, J.L. & Cantacessi, C. (2019) Dysbiosis associated with acute helminth infections in herbivorous youngstock – observations and implications. *Scientific Reports*, **9**, 11121.

Peachey, L.E., Jenkins, T.P. & Cantacessi, C. (2017) This gut ain't big enough for both of us. Or is it? helminth-microbiota interactions in veterinary species. *Trends in Parasitology*, **33**, 619–632.

Peachey, L.E., Molena, R.A., Jenkins, T.P., Cesare, A. Di, Traversa, D., Hodgkinson, J.E., *et al.* (2018) The relationships between faecal egg counts and gut microbial composition in UK Thoroughbreds infected by cyathostomins. *International Journal for Parasitology*, **48**, 403–412.

Pearson, G., Davies, L., White, A. & O'Brien, J. (1993) Pathological lesions associated with Anoplocephala perfoliata at the ileo-caecal junction of horses. *The Veterinary Record*, **132**, 179–182.

Penington, J.S., Penno, M.A.S., Ngui, K.M., Ajami, N.J., Roth-Schulze, A.J., Wilcox, S.A., *et al.* (2018) Influence of fecal collection conditions and 16S rRNA gene sequencing at two centers on human gut microbiota analysis. *Scientific Reports*, **8**, 1–10.

PerkinElmer. (2007) Clarus 600/560 D Gas Chromatograph/Mass Spectrometer (GC/MS) Tutorial. PerkinElmer, USA, p. https://www.perkinelmer.com/CMSResources/Images/44.

Perkins, G.A., Bakker, H.C. den, Burton, A.J., Erb, H.N., McDonough, S.P., McDonough, P.L., *et al.* (2012) Equine stomachs harbor an abundant and diverse mucosal microbiota. *Applied and Environmental Microbiology*, **78**, 2522–2532.

Perry, E., Cross, T.W.L., Francis, J.M., Holscher, H.D., Clark, S.D. & Swanson, K.S. (2018) Effect of road transport on the equine cecal microbiota. *Journal of Equine Veterinary Science*, **68**, 12–20.

Phua, L.C., Koh, P.K., Cheah, P.Y., Ho, H.K. & Eric, C.Y.C. (2013) Global gas chromatography / time-of-flight mass spectrometry (GC / TOFMS) -based metabonomic profiling of lyophilized human feces. *Journal of Chromatography B*, **937**, 103–113.

Pittaway, C.E., Lawson, A.L., Coles, G.C. & Douglas Wilson, A. (2014) Systemic and mucosal IgE antibody responses of horses to infection with Anoplocephala perfoliata. *Veterinary Parasitology*, **199**, 32–41.

Pleil, J.D., Stiegel, M.A. & Risby, T.H. (2013) Clinical breath analysis: Discriminating between human endogenous compounds and exogenous (environmental) chemical confounders. *Journal of Breath Research*, **7**.

Pollock, J., Glendinning, L., Wisedchanwet, T. & Watson, M. (2018) The madness of microbiome: Attempting to find consensus "best practice" for 16S microbiome studies. *Applied and Environmental Microbiology*, **84**, e02627-17.

Potter, T.L. & Fagerson, I.S. (1990) Composition of coriander leaf volatiles. *Journal of Agricultural and Food Chemistry*, **38**, 2054–2056.

Povolo, M. & Contarini, G. (2003) Comparison of solid-phase microextraxtion and purge and trap methods for the analysis of the volatile fraction butter. *Journal of Chromatography A*, **985**, 117–125.

Preter, V. De, Staeyen, G. Van, Esser, D., Rutgeerts, P. & Verbeke, K. (2009) Development of a screening method to determine the pattern of fermentation metabolites in faecal samples using on-line purge-and-trap gas chromatographic-mass spectrometric analysis. *Journal of chromatography. A*, **1216**, 1476–83.

Price, M.N., Dehal, P.S. & Arkin, A.P. (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, **5**, e9490.

Primec, M., Mičetić-Turk, D. & Langerholc, T. (2017) Analysis of short-chain fatty acids in human feces: A scoping review. *Analytical Biochemistry*, **526**, 9–21.

Proctor, G.B. & Carpenter, G.H. (2007) Regulation of salivary gland function by autonomic nerves. *Autonomic Neuroscience: Basic and Clinical*, **133**, 3–18.

Proudman, C. (2017) Intestinal Parasitism. In *The Equine Acute Abdomen* (ed. by Blikslager, A.T., White, N.A. & Moore, J.). Wiley-Blackwell, Oxford, pp. 193–203.

Proudman, C. & Edwards, G. (1992) Validation of a centrifugation/floatation technique for the diagnosis of equine cestodiasis. *The Veterinary Record*, **131**, 71–72.

Proudman, C., Holmes, M., Sheoran, A., Edwards, S. & Trees, A. (1997) Immunoepidemiology of the equine tapeworm Anoplocephala perfoliata: age-intensity profile and age-dependency of antibody subtype responses. *Parasitology*, **114**, 89–94.

Proudman, C. & Trees, A. (1996a) Use of excretory/secretory antigens for the serodiagnosis of Anoplocephala perfoliata cestodosis. *Veterinary Parasitology*, **61**, 239–247.

Proudman, C. & Trees, A. (1996b) Correlation of antigen specific IgG and IgG(T) responses with Anoplocephala perfoliata infection intensity in the horse. *Parasite immunology*, **18**, 499–506.

Proudman, C.J. (1994) The equine tapeworm. *Equine Veterinary Education*, **6**, 9–12.

Proudman, C.J. (2003) Diagnosis , Treatment , and Prevention of Tapeworm-associated Colic. *Journal of Equine Veterinary Science*, **23**, 6–9.

Proudman, C.J., French, N. & Trees, A. (1998) Tapeworm infection is a significant risk factor for spasmodic colic and ileal impaction colic in the horse. *Equine Veterinary Journal*, **30**, 194–199.

Proudman, C.J. & Holdstock, N.B. (2000) Investigation of an outbreak of tapeworm-associated colic in a training yard. *Equine Veterinary Journal*, **32**, 37–41.

Proudman, C.J., Hunter, J.O., Darby, a C., Escalona, E.E., Batty, C. & Turner, C. (2015) Characterisation of the faecal metabolome and microbiome of Thoroughbred racehorses. *Equine Veterinary Journal*, **47**, 580–586.

Purkhart, R., Köhler, H., Liebler-Tenorio, E., Meyer, M., Becher, G., Kikowatz, A., *et al.* (2011) Chronic intestinal Mycobacteria infection: Discrimination via VOC analysis in exhaled breath and headspace of feces using differential ion mobility spectrometry. *Journal of Breath Research*, **5**, 027103.

Putman, R., Pratt, R., Ekins, J. & Edwards, P. (1987) Food and Feeding Behaviour of Cattle and Ponies in the New Forest , Hampshire. *Journal of Applied Ecology*, **24**, 369–380.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, **41**, 590–596.

Quercia, S., Freccero, F., Castagnetti, C., Soverini, M., Turroni, S., Biagi, E., *et al.* (2019) Early colonisation and temporal dynamics of the gut microbial ecosystem in Standardbred foals. *Equine Veterinary Journal*, **51**, 231–237.

Raad, M. de, Fischer, C.R. & Northen, T.R. (2016a) High-throughput platforms for metabolomics. *Current Opinion in Chemical Biology*, **30**, 7–13.

Raad, M. de, Fischer, C.R. & Northen, T.R. (2016b) High-throughput platforms for metabolomics. *Current Opinion in Chemical Biology*, **30**, 7–13.

Ransom-Jones, E., Jones, D.L., McCarthy, A.J. & McDonald, J.E. (2012) The Fibrobacteres: an important phylum of cellulose-degrading bacteria. *Microbial ecology*, **63**, 267–81.

Rapin, A. & Harris, N.L. (2018) Helminth–bacterial interactions: cause and consequence. *Trends in Immunology*, **39**, 724–733.

Reade, S., Mayor, A., Aggio, R., Khalid, T., Pritchard, D., Ewer, A., *et al.* (2014) Optimisation of Sample Preparation for Direct SPME-GC-MS Analysis of Murine and Human Faecal Volatile Organic Compounds for Metabolomic Studies. *Journal of Analytical & Bioanalytical Techniques*, **5**, 184.

Reade, S., Williams, J.M., Aggio, R., Duckworth, C.A., Mahalhal, A., Hough, R., *et al.* (2019) Potential role of fecal volatile organic compounds as biomarkers of chemically induced intestinal inflammation in mice. *FASEB Journal*, **33**, 3129–3136.

Reardon, C., Sanchez, A.N.A., Hogaboam, C.M., Kay, D.M.M.C. & Weinstock, J. V. (2001) Tapeworm infection reduces epithelial ion transport abnormalities in murine dextran sulfate sodium-induced colitis. *Infection and Immunity*, **69**, 4417–4423.

Rehbein, S., Visser, M. & Winter, R. (2013) Prevalence, intensity and seasonality of gastrointestinal parasites in abattoir horses in Germany. *Parasitology Research*, **112**, 407–413.

Rehman, M.U., Zhang, H., Iqbal, M.K., Mehmood, K., Huang, S., Nabi, F., *et al.* (2017) Antibiotic resistance, serogroups, virulence genes, and phylogenetic groups of Escherichia coli isolated from yaks with diarrhea in Qinghai Plateau, China. *Gut Pathogens*, **9**, DOI 10.1186/s13099-017-0174-0.

Reinemeyer, C., Douglas, E., Wm, P., Marchiondo, A. & Jack, I. (2006) Dose-confirmation

studies of the cestocidal activity of pyrantel pamoate paste in horses. *Veterinary Parasitology*, **138**, 234–239.

Reinemeyer, C.R. & Nielsen, M.K. (2009) Parasitism and colic. *The Veterinary clinics of North America. Equine practice*, **25**, 233–45.

Rendle, D., Austin, C., Bowen, M., Cameron, I., Furtado, T., Hodgkinson, J., *et al.* (2019) Equine de-worming: a consensus on current best practice. *UK-Vet Equine*, **3**, 1–14.

Renker, C., Otto, P., Schneider, K., Zimdars, B., Maraun, M. & Buscot, F. (2005) Oribatid mites as potential vectors for soil microfungi: Study of mite-associated fungal species. *Microbial Ecology*, **50**, 518–528.

Resende Co, T., Hirsch, C., Toossi, Z., Dietze, R. & Ribeiro-Rodrigues, R. (2007) Intestinal helminth co-infection has a negative impact on both anti-Mycobacterium tuberculosis immunity and clinical response to tuberculosis therapy. *Clinical and Experimental Immunology*, **147**, 45–52.

Respondek, F., Goachet, a. G. & Julliand, V. (2008) Effects of dietary short-chain fructooligosaccharides on the intestinal microflora of horses subjected to a sudden change in diet. *Journal of Animal Science*, **86**, 316–323.

Reynolds, L. a, Smith, K. a, Filbey, K.J., Harcus, Y., Hewitson, J.P., Redpath, S. a, *et al.* (2014a) Commensal-pathogen interactions in the intestinal tract: lactobacilli promote infection with, and are promoted by, helminth parasites. *Gut microbes*, **5**, 522–32.

Reynolds, L.A., Harcus, Y., Smith, K.A., Webb, L.M., Hewitson, J.P., Ross, E.A., *et al.* (2014b) MyD88 Signaling Inhibits Protective Immunity to the Gastrointestinal Helminth Parasite Heligmosomoides polygyrus. *The Journal of Immunology*, **193**, 2984–2993.

Richard, M.L. & Sokol, H. (2019) The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases. *Nature Reviews Gastroenterology and Hepatology*, **16**, 331–345.

Richards, N., Choct, M., Hinch, G.N. & Rowe, J.B. (2004) Examination of the use of exogenous α-amylase and amyloglucosidase to enhance starch digestion in the small intestine of the horse. *Animal Feed Science and Technology*, **114**, 295–305.

Risticevic, S., Lord, H., Górecki, T., Arthur, C.L. & Pawliszyn, J. (2010) Protocol for solid-phase microextraction method development. *Nature protocols*, **5**, 122–39.

Rizzetto, L., Filippo, C. De & Cavalieri, D. (2014) Richness and diversity of mammalian fungal communities shape innate and adaptive immunity in health and disease. *European Journal of Immunology*, **44**, 3166–3181.

Roberts, D.D., Pollien, P. & Milo, C. (2000) Solid-Phase Microextraction Method Development for Headspace Analysis of Volatile Flavor Compounds. *Journal of Agricultural and Food Chemistry*, **48**, 2430–2437.

Roberts, L., Souza, A., Gerszten, R. & Clish, C. (2012) Targeted Metabolomics. *Current Protocols in Molecular Biology*, **CHAPTER:Un**, doi:10.1002/0471142727.mb3002s98.

Rodriguez, C., Taminiau, B., Brévers, B., Avesani, V., Broeck, J. Van, Leroux, A., *et al.* (2015) Faecal microbiota characterisation of horses using 16 rDNA barcoded pyrosequencing, and carriage rate of clostridium difficile at hospital admission. *BMC Microbiology*, **15**, https://doi.org/10.1186/s12866-015-0514-5.

Roesch, L.F.W., Casella, G., Simell, O., Krischer, J., Wasserfall, C.H., Schatz, D., *et al.* (2009)

Influence of fecal sample storage on bacterial community diversity. *The Open Microbiology Journal*, **3**, 40–46.

Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.A. (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology*, **13**, 1–19.

Round, J.L. & Mazmanian, S.K. (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nature Reviews Immunology*, **9**, 313–323.

Rowan, D.D. (2011) Volatile metabolites. *Metabolites*, **1**, 41–63.

Rubenstein, D.I. (1981) Behavioural ecology of island feral horses. *Equine Veterinary Journal*, **13**, 27–34.

Ryu, S., Bak, U., Kim, J., Y, H., Seo, H., Kim, J., *et al.* (2001) Cecal rupture by Anoplocephala perfoliata infection in a Thoroughbred horse in Seoul Race Park, South Korea. *Journal of Veterinary Science*, **2**, 189–193.

Sahuri-Arisoylu, M., Brody, L.P., Parkinson, J.R., Parkes, H., Navaratnam, N., Miller, A.D., *et al.* (2016) Reprogramming of hepatic fat accumulation and "browning" of adipose tissue by the short-chain fatty acid acetate. *International Journal of Obesity*, **40**, 955–963.

Salem, S. (2016) *Epidemiological and microbiome studies of equine colic*. Doctoral thesis, University of Liverpool.

Salem, S.E., Hough, R., Probert, C., Maddox, T.W., Antczak, P., Ketley, J.M., *et al.* (2019) A longitudinal study of the faecal microbiome and metabolome of periparturient mares. *PeerJ*, **7**, e6687.

Salem, S.E., Maddox, T.W., Berg, A., Antczak, P., Ketley, J.M., Williams, N.J., *et al.* (2018) Variation in faecal microbiota in a group of horses managed at pasture over a 12-month period. *Scientific Reports*, **8**, 8510.

Santos, A.S., Jerónimo, E., Ferreira, L.M., Rodrigues, M.A.M. & Bessa, R.J.B. (2010) Technical note : Fatty acids and purine profile of cecum and colon bacteria as indicators of equine microbial metabolism. *Journal of Animal Science*, **91**, 1753–1757.

Saric, J., Wang, Y., Li, J., Coen, M., Utzinger, J., Marchesi, J.R., *et al.* (2008) Species variation in the fecal metabolome gives insight into differential gastrointestinal function. *Journal of Proteome Research*, **7**, 352–360.

Sarkijarvi, S., Niemelainen, O., Sormunen-Cristian, R. & Saastamoinen, M. (2012) Changes in chemical composition of different grass species and -mixtures in equine pasture during grazing season. In *Forages and Grazing in horse nutrition* (ed. by Saastamoinen, M., Fradinho, M., Santos, A. & Miraglia, N.). Wageningen Academic Publishers, pp. 45–48.

Savage, D., Siegel, J., Snellen, J. & Whitt, D. (1981) Transit time of epithelial cells in the small intestine of germfree mice and ex-germfree mice associated with indigenous microorganisms. *Applied Environmental Microbiology*, **42**, 996–1001.

Scher, J.U., Sczesnak, A., Longman, R.S., Segata, N., Ubeda, C., Bielski, C., *et al.* (2013) Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *eLife*, **2**, 1–20.

Schirmer, M., Franzosa, E.A., Lloyd-Price, J., McIver, L.J., Schwager, R., Poon, T.W., *et al.* (2018) Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nature Microbiology*, **3**, 337–346.

Schoster, A., Weese, J.S. & Guardabassi, L. (2014) Probiotic use in horses - what is the evidence for their clinical efficacy? *Journal of veterinary internal medicine / American College of Veterinary Internal Medicine*, **28**, 1640–52.

Semren, T.Ž., Karačonji, I.B., Safner, T., Brajenović, N., Lovaković, B.T. & Pizent, A. (2018) Gas chromatographic-mass spectrometric analysis of urinary volatile organic metabolites: optimization of the HS-SPME procedure and sample storage conditions. *Talanta*, **176**, 537–543.

Sheikha, A. El, Durand, N., Sarter, S., Okullo, J. & Montet, C. (2012) Study of the microbial discrimination of fruits by PCR-DGGE: application to the determination of the geographical origin of Physalis fruits from Columbia, Egypt, Uganda and Madagascar. *Food Control*, **24**, 57–63.

Shepherd, M.L., Ponder, M. a, Burk, A.O., Milton, S.C. & Swecker, W.S. (2014) Fibre digestibility, abundance of faecal bacteria and plasma acetate concentrations in overweight adult mares. *Journal of Nutritional Science*, **3**, e10.

Shepherd, M.L., Swecker, W.S., Jensen, R. V. & Ponder, M.A. (2012) Characterization of the fecal bacteria communities of forage-fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. *FEMS Microbiology Letters*, **326**, 62–68.

Shepherd, M.L., Swecker, W.S. & Ponder, M.A. (2015) Effect of two different commercial DNA extraction kits on the bacterial 16S ribosomal RNA gene denaturing gradient gel electrophoresis profile of Arabian gelding feces. *Journal of Equine Veterinary Science*, **35**, 165–169.

Shirazi-Beechey, S.P. (2008) Molecular insights into dietary induced colic in the horse. *Equine veterinary journal*, **40**, 414–21.

Shirey, R.E. (2012) Handbook of Solid Phase Microextraction. In *Handbook of Solid Phase Microextraction* (ed. by Pawliszyn, J.). Elsevier, pp. 99–133.

Shurin, J. (2000) Dispersal limitation, invasion resistance, and the structure of pond zooplankton communities. *Ecology*, **8**, 3074–3086.

Simmons, H. & Ford, E. (1991) Gluconeogenesis from propionate produced in the colon of the horse. *British Veterinary Journal*, **147**, 340–345.

Simpson, J.M., Martineau, B., Jones, W.E., Ballam, J.M. & Mackie, R.I. (2002) Characterization of fecal bacterial populations in canines: effects of age, breed and dietary fiber. *Microbial Ecology*, **44**, 186–197.

Sipos, R., Székely, A.J., Palatinszky, M., Révész, S., Márialigeti, K. & Nikolausz, M. (2007) Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiology Ecology*, **60**, 341–350.

Skotarek, S.L., Colwell, D.D. & Goater, C.P. (2010) Evaluation of diagnostic techniques for Anoplocephala perfoliata in horses from Alberta, Canada. *Veterinary Parasitology*, **172**, 249–255.

Šlapeta, J., Dowd, S.E., Alanazi, A.D., Westman, M.E. & Brown, G.K. (2015) Differences in the faecal microbiome of non-diarrhoeic clinically healthy dogs and cats associated with Giardia duodenalis infection: Impact of hookworms and coccidia. *International Journal for Parasitology*, **45**, 585–594.

Snalune, K., Hunter, J., Waring, R., Turner C & Batty C. (2019) Modulation of the equine

microbiome by pasture and feed supplements: A metabolomics approach. *Integrated Food, Nutrition and Metabolism*, **6**, 1–4.

Sneddon, J.C. & Argenzio, R.A. (1998) Feeding strategy and water homeostasis in equids: The role of the hind gut. *Journal of Arid Environments*, **38**, 493–509.

Snelling, T. (2013) *Cellulose and hemicellulose degradation by cilliates from the caecum of the horse*. Doctoral thesis, Aberystwyth University.

Snelling, T.J., Pinloche, E., Worgan, H., Newbold, C. & Mcewan, N.R. (2011) Molecular phylogeny of Spirodinium equi, Triadinium caudatum and Blepharocorys sp. from the equine hindgut. *Acta Protozoologica*, **50**, 319–326.

Snyder, L.R. & Kirkland, J.J. (1979) Basic Concepts and Control of Separation. In *Introduction to Modern Liquid Chromatography*. Wiley-Interscience, Toronto, pp. 15–82.

Spiljar, M., Merkler, D. & Trajkovski, M. (2017) The immune system bridges the gut microbiota with systemic energy homeostasis: Focus on TLRs, mucosal barrier, and SCFAs. *Frontiers in Immunology*, **8**, 1–10.

St-Pierre, B., la Fuente, G. de, O'Neill, S., Wright, A.-D.G. & Jassim, R. Al. (2013) Analysis of stomach bacterial communities in Australian feral horses. *Molecular Biology Reports*, **40**, 369–76.

Stange, E.F. & Schroeder, B.O. (2019) Microbiota and mucosal defense in IBD: an update. *Expert Review of Gastroenterology and Hepatology*, **13**, 963–976.

Stavert, J.R., Drayton, B., Beggs, J.R. & Gaskett, A.C. (2014) The volatile organic compounds of introduced and native dung and carrion and their role in dung beetle foraging behaviour. *Ecological Entomology*, **39**, 556–565.

Steelman, S.M., Chowdhary, B.P., Dowd, S., Suchodolski, J. & Janečka, J.E. (2012) Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC veterinary research*, **8**, 231.

Stevens, C.E. & Hume, I.D. (1998) Contributions of microbes in vertebrate gastrointestinal tract to production and conservation of nutrients. *Physiological reviews*, **78**, 393–427.

Su Yin Ng, J., Ryan, U., Trengove, R. & Maker, G. (2012) Development of an untargeted metabolomics method for the anlysis of human faecal samples using Cryptosporidium-infected samples. *Molecular & Biochemical Parasitology*, **185**, 145–150.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., *et al.* (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics*, **3**, 211–221.

Suthers, J., Vale, L. & Archer, D. (2014) Prevalence of different types of colic in broodamares and short term-survival in a UK equine referral hospital population: 146 cases (2000-2012). In *Eleventh Equine Colic Research Symposium*. p. https://docplayer.net/13449869-Abstract-programme-.

Suthers, J.M., Pinchbeck, G.L., Proudman, C.J. & Archer, D.C. (2013) Risk factors for large colon volvulus in the UK. *Equine veterinary journal*, **45**, 558–63.

Tang, Z.-Z., Chen, G., Hong, Q., Huang, S., Smith, H.M., Shah, R.D., *et al.* (2019) Multi-omic analysis of the microbiome and metabolome in healthy subjects reveals microbiome-dependent relationships between diet and metabolites. *Frontiers in Genetics*, **10**.

Tapio, I., Fischer, D., Blasco, L., Tapio, M., Wallace, R.J., Bayat, A.R., *et al.* (2017) Taxon abundance, diversity, co-occurrence and network analysis of the ruminal microbiota in response to dietary changes in dairy cows. *PLoS ONE*, **12**, 1–21.

Tardy, F., Gaurivaud, P., Tricot, A., Maigre, L. & Poumarat, F. (2009) Epidemiological surveillance of mycoplasmas belonging to the ' Mycoplasma mycoides ' cluster : is DGGE fingerprinting of 16S rRNA genes suitable ? *Letters in Applied Microbiology*, **48**, 210–217.

Tateo, A., Padalino, B., Boccaccio, M., Maggiolino, A. & Centoducati, P. (2012) Transport stress in horses: Effects of two different distances. *Journal of Veterinary Behavior: Clinical Applications and Research*, **7**, 33–42.

Tomczuk, K., Kostro, K., Grzybek, M., Szczepaniak, K., Studzińska, M., Demkowska-Kutrzepa, M., *et al.* (2015) Seasonal changes of diagnostic potential in the detection of Anoplocephala perfoliata equine infections in the climate of Central Europe. *Parasitology Research*, **114**, 767–772.

Tomczuk, K., Kostro, K., Szczepaniak, K.O., Grzybek, M., Studzińska, M., Demkowska-Kutrzepa, M., *et al.* (2014) Comparison of the sensitivity of coprological methods in detecting Anoplocephala perfoliata invasions. *Parasitology research*, **113**, 2401–6.

Traub-Dargatz, J.L., Kopral, C.A., Seitzinger, a H., Garber, L.P., Forde, K. & White, N.A. (2001) Estimate of the national incidence of and operation-level risk factors for colic among horses in the United States, spring 1998 to spring 1999. *Journal of the American Veterinary Medical Association*, **219**, 67–71.

Trotz-Williams, L., Physick-Sheard, P., McFarlane, H., Pearl, D.L., Martin, S.W. & Peregrine, A.S. (2008) Occurrence of Anoplocephala perfoliata infection in horses in Ontario, Canada and associations with colic and management practices. *Veterinary parasitology*, **153**, 73–84.

Tuckwell, D.S., Nicholson, M.J., McSweeney, C.S., Theodorou, M.K. & Brookman, J.L. (2005) The rapid assignment of ruminal fungi to presumptive genera using ITS1 and ITS2 RNA secondary structures to produce group-specific fingerprints. *Microbiology*, **151**, 1557–1567.

Turnbaugh, P.J., Bäckhed, F., Fulton, L. & Gordon, J.I. (2008) Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host and Microbe*, **3**, 213–223.

Turner, C. (2016) Techniques and issues in breath and clinical sample headspace analysis for disease diagnosis. *Bioanalysis*, **8**, 677–690.

Turner, C., Batty, C., Escalona, E., Hunter, J. & Proudman, C. (2013) The Use of SIFT-MS in Profiling the Faecal Volatile Metabolome in Horses with Colic: A Pilot Study. *Current Analytical Chemistry*, **9**, 614–621.

Uden, P. & Soest, P. Van. (1982) Comparative digestion of timothy (Phleum pratense) fibre by ruminants, equines and rabbits. *The British Journal of Nutrition*, **47**, 267–272.

Vandenbergh, P. a. (1993) Lactic acid bacteria, their metabolic products and interference with microbial growth. *FEMS Microbiology Reviews*, **12 (1–3)**, 221–237.

Venable, E., Reed, S., Jacobs, R., Jerina, M. & ME, G. (2017a) Development and management of a cannulated horse herd. *Journal of Equine Veterinary Science*, **52**, 55–56.

Venable, E.B., Fenton, K.A., Braner, V.M., Reddington, C.E., Halpin, M.J., Heitz, S.A., *et al.* (2017b) Effects of feeding management on the equine cecal microbiota. *Journal of Equine Veterinary Science*, **49**, 113–121.

Vermorel, M. (1997) Yearly methane emsiions of digestive origin by sheep, goats and equines in France. Variations with physiological stage and productive type. *INRA Productions Animales*, **10**, 153–161.

Wagget, B., McGorum, B., Wemery, U., Shaw, D. & Pirie, R. (2010) Prevalence of Clostridium perfringens in faeces and ileal contents from grass sickness affected horses: comparison with 3 control populations. *Equine Veterinary Journal*, **42**, 494–499.

Walk, S.T., Blum, A.M., Ewing, S.A.-S., Weinstock, J. V & Young, V.B. (2010) Alteration of the murine gut microbiota during infection with the parasitic helminth Heligmosomoides polygyrus. *Inflammatory bowel diseases*, **16**, 1841–9.

Waring, R.H., Hunter, J.O., Turner, C., Batty, C. & Ramzan, P.H.L. (2019) Nitrate supplementation in thoroughbred racehorses : Addition of beetroot juice to the equine diet and effects on the gut metabolome. *Integrated Food, Nutrition and Metabolism*, **6**, doi: 10.15761/IFNM.1000246.

Weese, J. (2003) Probiotics, Prebiotics and Synbiotics. In *Current Therapy in Equine Medicine* (ed. by Sprayberry, K. & Robinson, N.). pp. 711–714.

Weese, J., Staempfli, H.R. & Prescott, J. (2001) A prospective study of the roles of Clostridium difficile and enterotoxigenic Clostridium perfringens in equine diarrhoea. *Equine Veterinary Journal*, **33**, 403–409.

Weese, J.S., Anderson, M.E.C., Lowe, A., Penno, R., Costa, T.M. da, Button, L., *et al.* (2004) Screening of the equine intestinal microflora for potential probiotic organisms. *Equine Veterinary Journal*, **36**, 351–355.

Weese, J.S., Holcombe, S.J., Embertson, R.M., Kurtz, K. a, Roessner, H. a, Jalali, M., *et al.* (2014) Changes in the faecal microbiota of mares precede the development of post partum colic. *Equine Veterinary Journal*, **47**, 641–649.

Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., *et al.* (2018) Missing value imputation approach for mass spectrometry-based metabolomics data. *Scientific Reports*, **8**, 1–10.

Wen, H., Lu, C., Yuan, Z., Wang, X. & Su, S. (2018) Analysis of gut fungal community of cows with clinical mastitis. *Advances in Microbiology*, **08**, 366–377.

West, J.S., Townsend, J.A., Stevens, M. & Fitt, B.D.L. (2012) Comparative biology of different plant pathogens to estimate effects of climate change on crop diseases in Europe. *European Journal of Plant Pathology*, **133**, 315–331.

Weyenberg, S. Van, Sales, J. & Janssens, G. (2006) Passage rate of digesta through the equine gastrointestinal tract: a review. *Livestock Science*, **99**, 3–12.

Wheatley, R.E., Millar, S.E. & Griffiths, D.W. (1996) The production of volatile organic compounds during nitrogen transformations in soils. *Plant and Soil*, **181**, 163–167.

Whitfield-Cargile, C.M., Cohen, N.D., Suchodolski, J., Chaffin, M.K., McQueen, C.M., Arnold, C.E., *et al.* (2015) Composition and diversity of the fecal microbiome and inferred fecal metagenome does not predict subsequent pneumonia caused by rhodococcus equi in foals. *PLoS ONE*, **10**, 1–19.

Williams, S., Horner, J., Orton, E., Green, M., McMullen, S., Mobasheri, a, *et al.* (2015) Water intake, faecal output and intestinal motility in horses moved from pasture to a stabled management regime with controlled exercise. *Equine veterinary journal*, **47**, 96–100.

Williamson, R.M.C., Gasser, R.B., Middleton, D. & Beveridge, I. (1997) The distribution of Anoplocephala perfoliata in the intestine of the horse and associated pathological changes. *Veterinary Parasitology*, **73**, 225–241.

Willing, B., Vörös, a., Roos, S., Jones, C., Jansson, a. & Lindberg, J.E. (2009) Changes in faecal bacteria associated with concentrate and forage-only diets fed to horses in training. *Equine Veterinary Journal*, **41**, 908–914.

Windey, K., Preter, V. de & Verbeke, K. (2012) Relevance of protein fermentation to gut health. *Molecular Nutrition and Food Research*, **56**, 184–196.

Wong, K., Shaw, T.I., Oladeinde, A., Glenn, T.C., Oakley, B. & Molina, M. (2016) Rapid microbiome changes in freshly deposited cow feces under field conditions. *Frontiers in Microbiology*, **7**, doi: 10.3389/fmicb.2016.00500.

Wood, D.E.L., Matthews, J.B., Stephenson, S., Slote, M. & Nussey, D.H. (2013) Variation in fecal egg counts in horses managed for conservation purposes: Individual egg shedding consistency, age effects and seasonal variation. *Parasitology*, **140**, 115–128.

Woolfenden, E. (2010) Sorbent-based sampling methods for volatile and semi-volatile organic compounds in air. Part 2. Sorbent selection and other aspects of optimizing air monitoring methods. *Journal of Chromatography A*, **1217**, 2685–2694.

Woolfenden, E. (2012) Thermal desorption for gas chromatography. In *Gas Chromatography* (ed. by Poole, C.). Elsevier, pp. 235–290.

Wu, S., Li, R.W., Li, W., Beshah, E., Dawson, H.D. & Urban, J.F. (2012) Worm burden-dependent disruption of the porcine colon microbiota by Trichuris suis infection. *PloS one*, **7**, e35470.

Wyk, J.A. Van. (2001) Refugia--overlooked as perhaps the most potent factor concerning the development of anthelmintic resistance. *The Onderstepoort Journal of Veterinary Research*, **68**, 55–67.

Wylie, C.E., Shaw, D.J., Fordyce, F.M., Lilly, A., Pirie, R.S. & McGorum, B.C. (2016) Equine grass sickness in Scotland: A case-control study of environmental geochemical risk factors. *Equine Veterinary Journal*, **48**, 779–785.

Wypych, G. (2017) Odor in various products. In *Handbook of Odors in Plastic Materials* (ed. by Wypych, G.). Chem Tec Publishing, Toronto, Canada, pp. 101–137.

Xia, J., Broadhurst, D.I., Wilson, M. & Wishart, D.S. (2013) Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics*, **9**, 280–299.

Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D. & Wishart, D.S. (2012) MetaboAnalyst 2.0-a comprehensive server for metabolomic data analysis. *Nucleic Acids Research*, **40**, 127–133.

Yamano, H., Koike, S., Kobayashi, Y. & Hata, H. (2008) Phylogenetic analysis of hindgut microbiota in Hokkaido native horses compared to light horses. *Animal Science Journal*, **79**, 234–242.

Yue, Q., Wang, C., Gianfagna, T.J. & Meyer, W.A. (2001) Volatile compounds of endophyte-free and infected tall fescue (Festuca arundinacea Schreb.). *Phytochemistry*, **58**, 935–941.

Yuki, N., Shimazaki, T., Kushiro, a., Watanabe, K., Uchida, K., Yuyama, T., *et al.* (2000) Colonization of the Stratified Squamous Epithelium of the Nonsecreting Area of Horse Stomach by Lactobacilli. *Applied and Environmental Microbiology*, **66**, 5030–5034.

Yuyama, T., Takai, S., Tsubaki, S., Kado, Y. & M, M. (2004) Evaluation of a host-specific Lactobacillus probiotics in training horses and neonatal foals. *Journal of Intestinal Microbiology*, **18**, 101–106.

Zaiss, M.M. & Harris, N.L. (2016) Interactions between the intestinal microbiome and helminth parasites. *Parasite Immunology*, **38**, 5–11.

Zaiss, M.M., Rapin, A., Lebon, L., Dubey, L.K., Mosconi, I., Sarter, K., *et al.* (2015) The intestinal microbiota contributes to the ability of helminths to modulate allergic inflammation. *Immunity*, **43**, 998–1010.

Zhang, I., Pletcher, S.D., Goldberg, A.N., Barker, B.M. & Cope, E.K. (2017) Fungal microbiota in chronic airway inflammatory disease and emerging relationships with the host immune response. *Frontiers in Microbiology*, **8**, 1–7.

Zhang, Z. & Li, G. (2010) A review of advances and new developments in the analysis of biological volatile organic compounds. *Microchemical Journal*, **95**, 127–139.

Zhao, S., Zhao, J., Bu, D., Sun, P., Wang, J. & Dong, Z. (2014) Metabolomics analysis reveals large effect of roughage types on rumen microbial metabolic profile in dairy cows. *Letters in Applied Microbiology*, **59**, 79–85.

Zhao, Y., Li, B., Bai, D., Huang, J., Shiraigo, W., Yang, L., *et al.* (2016) Comparison of the fecal microbiota of Mongolian and Thoroughbred horses by high-throughput sequencing of the V4 region of the 16S rRNA gene. *Asian-Australasian Journal of Animal Sciences*, **29**, 1345–1352.

Zoetendal, E.G., Collier, C.T., Koike, S., Mackie, R.I. & Gaskins, H.R. (2004) Molecular ecological analysis of the gastrointestinal microbiota : a review. *Journal of Nutrition*, **134**, 465–472.

Appendix

Appendix 2.1 List of VOCs and their abundances that were exclusive to 1000 and 2000 mg samples.

	Mean log abundance	
VOC	**1000 mg**	**2000 mg**
2-Butanone	16.43	16.36
Ethyl Acetate	17.07	17.09
n-Propyl acetate	15.6	15.57
Propanoic acid, 2-methyl-, ethyl ester	15.77	15.6
2-Hexanone	15.69	15.83
Butanoic acid, 2-methyl-, ethyl ester	16.22	16.19
2-Octene, 2,6-dimethyl-	15.64	15.55
Benzene, propyl-	15.86	15.8
2-Heptanone, 5-methyl-	17.77	17.64
Benzene, 1-methyl-3-propyl-	16.22	16.16
alpha-copaene	14.09	14.21
2-Undecenal	15.94	15.57
Acetone	16.11	16.41

Appendix 2.2 A list of principal component scores for sample mass.

VOC	**PC1**	**PC2**
x39.85_Pentadecane	0.173722	-0.0197
x31.43_Undecane..2.6.dimethyl.	0.172827	0.009826
x36.89_Tetradecane	0.172217	0.035548
x36.30_Dodecane..2.6.10.trimethyl.	0.169194	-0.0276
x27.15_Benzene..1.methyl.3.propyl.	0.168736	-0.02803
x20.60_Benzene..1.3.dimethyl.	0.164393	0.025442
x21.71_Styrene	0.16428	0.047629
x23.84_Benzene..propyl.	0.164233	-0.02295
x20.30_Ethylbenzene	0.163964	-0.01239
x34.03_Tridecane	0.163565	0.031892
x38.91_Tetradecane..3.methyl.	0.161863	-0.01465
x36.01_Tridecane..3.methyl.	0.159769	-0.01361
x25.44_Benzene..1.2.4.trimethyl.	0.159373	-0.04796
x18.25_Hexanal	0.154757	0.003279
x32.86_Dodecane..2.methyl.	0.15262	0.018229
x24.37_2.Heptanone..6.methyl.	0.151903	0.072944
x26.20_D.Limonene	0.150838	0.060581
x25.55_5.Hepten.2.one..6.methyl.	0.149511	0.075669
x20.49_Nonane	0.146092	-0.01108
x21.94_2.Heptanone	0.145826	0.062927
x25.66_2.Octanone	0.142379	0.000617
x38.70_Heptadecane..2.6.10.14.tetramethyl.	0.141883	0.096729

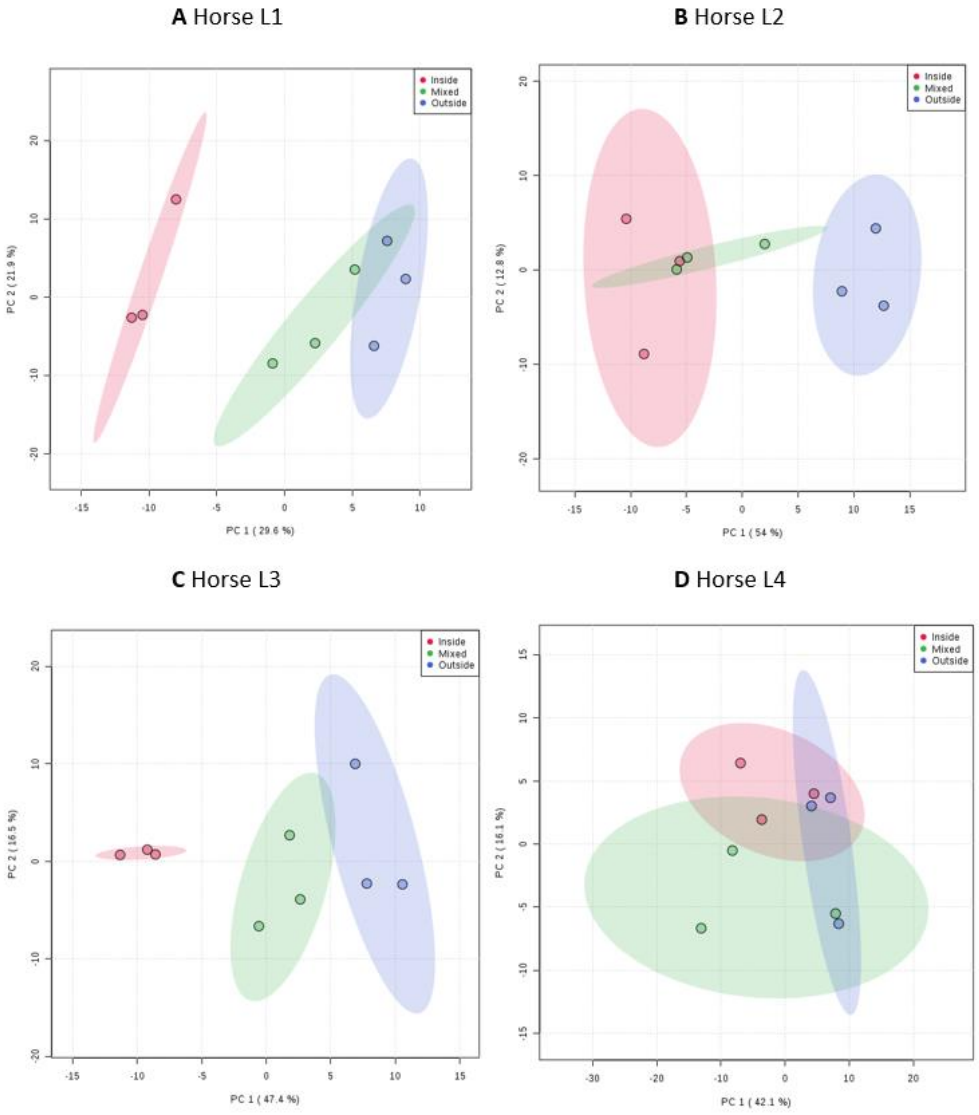
x24.21_Decane	0.141839	0.021958
x24.74_Furan..2.pentyl.	0.141518	0.083753
x39.02_Bicyclo.3.1.1.hept.2.ene..2.6.dimethyl.6..4.methyl.3.pentenyl..	0.140387	0.009916
x37.58_alpha.Copaene	0.136494	0.001576
x37.80_2.Undecenal	0.135944	0.004804
x24.81_2.Heptenal...Z..	0.131939	0.000606
x26.01_Octanal	0.131608	0.106153
x22.26_Heptanal	0.126055	0.093951
x29.51_Nonanal	0.125577	0.119273
x29.16_2.Nonanone	0.121338	0.106965
x35.85_Tridecane..2.methyl.	0.120883	0.026982
x29.99_Undecane..2.methyl.	0.116811	-0.06274
x27.70_Undecane	0.11055	0.019313
x29.00_Propanoic.acid..hexyl.ester	0.106966	0.008295
x16.66_Octane	0.105556	-0.03506
x32.41_2.Decanone	0.103372	0.090506
x16.93_1.Propanone..1.cyclopropyl.	0.094728	-0.04247
x12.79_Butanal..2.methyl.	0.092666	0.153499
x25.36_3.Octanone	0.086832	3.63E-05
x14.21_Pentanal	0.070757	0.088891
x35.46_2.Undecanone	0.067324	0.066628
x24.68_2.Heptanone..5.methyl.	0.066862	-0.00241
x25.441_Benzaldehyde	0.066813	0.137546
x27.66_Cyclohexanone..2.2.6.trimethyl.	0.065959	0.127756
x29.81_Decane..2.9.dimethyl.	0.063552	0.074068
x12.45_Butanal..3.methyl.	0.062443	0.1367
x17.61_Butanoic.acid..ethyl.ester	0.059864	0.114982
x32.44_Tetradecanal	0.055293	0.025667
x8.98_Propanal..2.methyl.	0.052702	0.107531
x21.00_2.Hexenal	0.052106	0.080677
x25.16_n.Caproic.acid.vinyl.ester	0.049554	0.152131
x26.53_Dimethyl.sulfone	0.032677	-0.13358
x32.76_Decanal	0.032241	0.095404
x25.33_1.Butanol..3.methyl...propanoate	0.032025	0.125566
x23.27_2.Octene..2.6.dimethyl.	0.027557	0.111384
x34.15_1.Cyclohexene.1.carboxaldehyde..2.6.6.trimethyl.	0.025172	0.122691
x7.35_Acetone	0.016915	0.032306
x10.50_2.Butanone	0.008593	0.002621
x18.01_2.Hexanone	0.005161	0.007708
x6.62_Ethanol	0.003365	0.020103
x22.05_1.7.Octadiene..2.7.dimethyl.	2.37E-05	0.075398
x24.41_beta.Pinene	-0.00058	0.116977
x35.73_1H.Pyrrole.2.5.dione..3.ethyl.4.methyl.	-0.00314	0.113444
x18.64_Methyl.valerate	-0.00582	0.149576
x14.10_Propanoic.acid..ethyl.ester	-0.01019	0.178058

x28.00_Phenol	-0.0116	0.146801
x24.12_cis.2.6.Dimethyl.2.6.octadiene	-0.01413	0.127973
x30.91_p.Cresol	-0.01589	0.106882
x28.64_Benzeneacetaldehyde	-0.01731	0.13357
x26.84_Benzene...1.methylethyl..	-0.01925	0.06802
x18.36_Propanoic.acid..2.methyl.	-0.01949	-0.03477
x12.55_Acetic.acid	-0.01958	-0.02617
x16.07_Propanoic.acid	-0.0255	0.020069
x22.52_1.6.Octadiene..3.7.dimethyl....S..	-0.0262	0.139162
x27.14_Butanoic.acid..2.methylbutyl.ester	-0.02799	0.167857
x16.68_Butanoic.acid..2.methyl...methyl.ester	-0.03319	0.159555
x11.15_Methyl.propionate	-0.03385	0.168875
x22.97_Pentanoic.acid	-0.03662	0.066692
x19.40_Butanoic.acid	-0.03752	0.03417
x20.17_Propanoic.acid..2.methylpropyl.ester	-0.0398	0.153969
x17.99_Propanoic.acid..propyl.ester	-0.03984	0.154196
x19.67_Butanoic.acid..3.methyl...ethyl.ester	-0.04061	0.164682
x14.30_n.Propyl.acetate	-0.04334	0.152304
x19.45_Butanoic.acid..2.methyl...ethyl.ester	-0.04343	0.152364
x15.91_Propanoic.acid..2.methyl...ethyl.ester	-0.04556	0.153236
x21.73_Butanoic.acid..2.methyl.	-0.05019	0.055304
x21.44_Butanoic.acid..3.methyl.	-0.05136	0.060247
x14.63_Butanoic.acid..methyl.ester	-0.05287	0.133918
x23.22_Butanoic.acid..3.methyl...propyl.ester	-0.05653	0.07819
x10.60_Ethyl.Acetate	-0.05801	0.157723
x30.32_Pentanoic.acid..3.methylbutyl.ester	-0.05854	0.155409
x31.50_Benzyl.methyl.ketone	-0.06403	0.149108
x16.92_2.Octene...E..	-0.06845	0.064401
x23.58_Butanoic.acid..2.methylpropyl.ester	-0.07376	0.158793
x28.27_Pentanoic.acid..butyl.ester	-0.07479	0.155674
x23.51_Methional	-0.07957	0.118509
x37.19_2.Hexadecene..3.7.11.15.tetramethyl....R..R..R...E....	-0.08671	-0.03281
x30.37_2.3H..Furanone..5.ethyldihydro.	-0.08771	0.141794
x19.13_Butanoic.acid..1.methylethyl.ester	-0.08872	0.15903
x24.46_2.6.Octadiene..2.6.dimethyl.	-0.0888	0.10515
x16.50_Toluene	-0.0945	0.141497
x38.57_Indole	-0.109	0.099126

Appendix 2.3 List of compounds exclusive to each SPME fibre.

DVB-CAR-PDMS	CAR-PDMS
2,3-Butanedione	beta-phellandrene
Furan, 2-ethyl-	
2-Pentanone	
Octane	
Propanoic acid, 2-methyl-	
Methional	
beta-pinene	
2-Heptanone, 5-methyl-	
2-Heptenal, (Z)-	
1-Hepten-3-one	
3-Octanone	
Dimethyl sulfone	
Benzene, (1-methylethyl)-	
Benzene, 1-methyl-3-propyl-	
Benzeneacetaldehyde	
2-Nonanone	
2(3H)-Furanone, 5-ethyldihydro-	
Benzyl methyl ketone	
2-Decanone	
1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	
Indole	

Appendix 2.4 PCA plots of the VOC profile inside, outside and mixed contents of the faecal balls of four horses.



Appendix 2.5 List of principal component scores for time lapse to freezing.

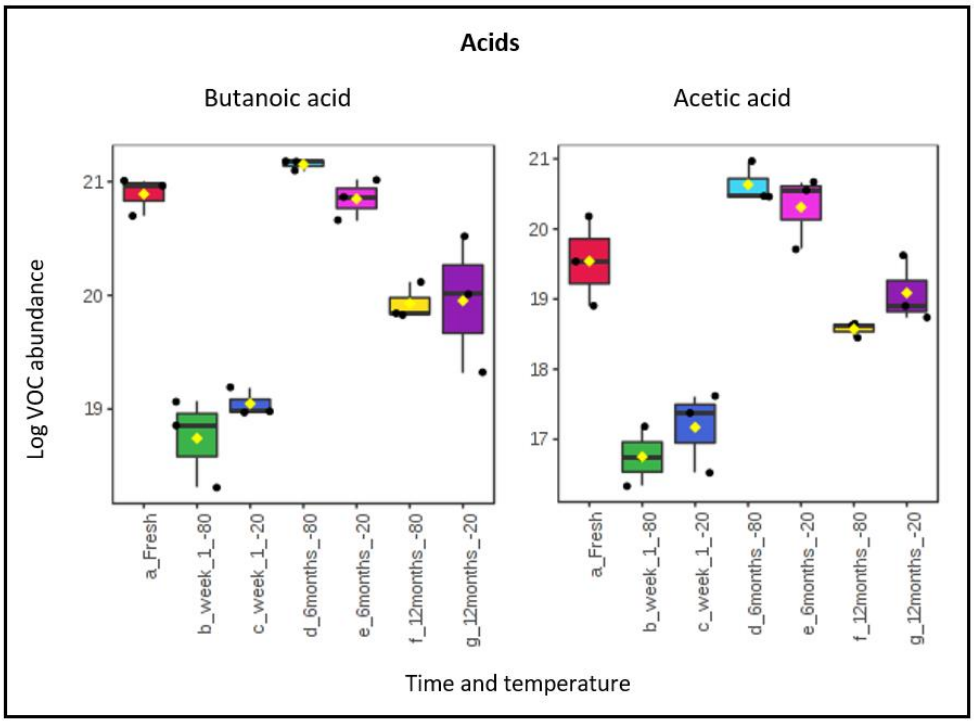
VOC	PC1	PC2
x10.22_2.3.Butanedione	0.024887	-0.15199
x10.50_2.Butanone	0.14419	-0.08666
x10.60_Ethyl.Acetate	-0.07609	-0.06958
x10.80_2.Butanol	-0.03217	-0.11551
x11.15_Methyl.propionate	-0.06868	-0.06447
x12.45_Butanal..3.methyl.	0.104435	-0.08727
x12.55_Acetic.acid	-0.06588	-0.02111
x12.79_Butanal..2.methyl.	0.136534	-0.06149
x13.05_Propanoic.acid..2.methyl...methyl.ester	-0.0921	-0.1477
x13.43_Furan..2.ethyl.	0.169799	-0.03719
x13.80_1.Penten.3.one	0.124331	0.011149

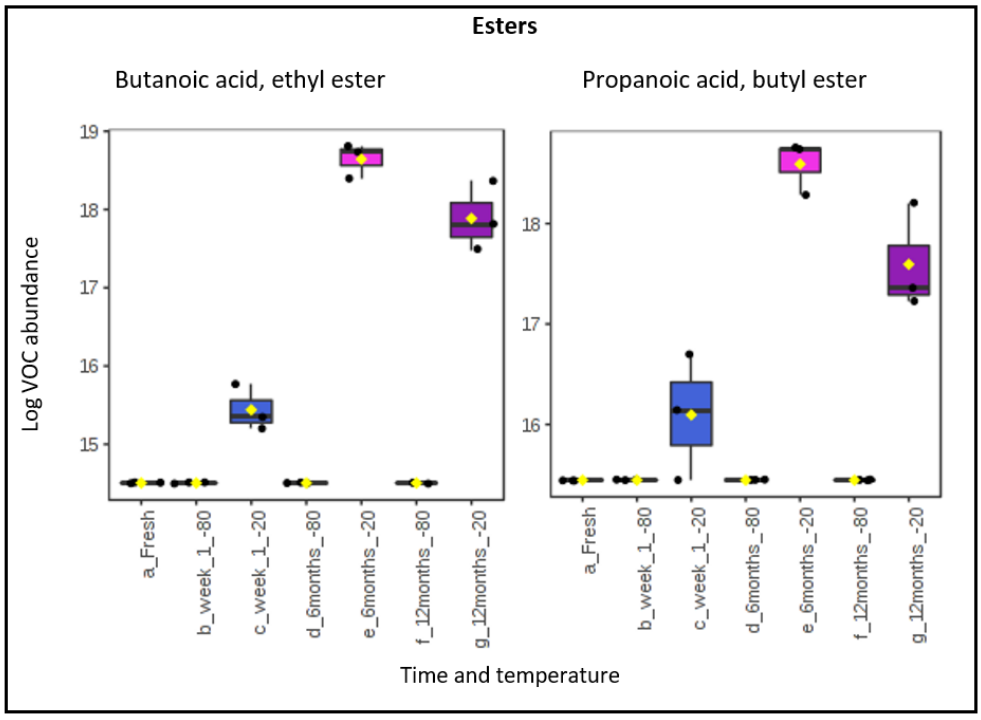
x13.88_2.Pentanone	0.153241	-0.08915
x14.10_Propanoic.acid..ethyl.ester	-0.14602	-0.06272
x14.21_Pentanal	0.138235	-0.07434
x14.30_n.Propyl.acetate	-0.11814	-0.07161
x14.63_Butanoic.acid..methyl.ester	-0.1297	-0.05487
x15.21_Heptane..2.methyl.	-0.04593	-0.0907
x15.79_1.Penten.3.one..4.methyl.	0.081584	0.046658
x15.91_Propanoic.acid..2.methyl...ethyl.ester	-0.07483	-0.07062
x16.07_Propanoic.acid	-0.03017	-0.12966
x16.19_3.Pentanone..2.methyl.	0.145646	-0.07662
x16.50_Toluene	-0.08588	-0.14525
x16.66_Octane	-0.00346	-0.18878
x16.68_Butanoic.acid..2.methyl...methyl.ester	0.001535	-0.04517
x16.93_1.Propanone..1.cyclopropyl.	0.170122	-0.02753
x17.61_Butanoic.acid..ethyl.ester	-0.04725	-0.12924
x17.99_Propanoic.acid..propyl.ester	-0.15483	-0.03977
x18.01_2.Hexanone	0.136695	-0.11726
x18.25_Hexanal	0.146877	-0.04364
x18.36_Propanoic.acid..2.methyl.	-0.06078	-0.1373
x18.64_Methyl.valerate	-0.14618	0.010815
x19.13_Butanoic.acid..1.methylethyl.ester	-0.01748	-0.07611
x19.40_Butanoic.acid	-0.03315	-0.07342
x19.65_Propanoic.acid..2.methyl...propyl.ester	-0.08866	0.006879
x20.30_Ethylbenzene	-0.01468	-0.11781
x20.49_Nonane	-0.07709	-0.14363
x20.60_Benzene..1.3.dimethyl.	0.00807	-0.15148
x21.00_2.Hexenal	0.150459	-0.10287
x21.44_Butanoic.acid..3.methyl.	-0.08141	-0.11111
x21.73_Butanoic.acid..2.methyl.	-0.09163	-0.1062
x21.94_2.Heptanone	0.111567	-0.13344
x22.05_1.7.Octadiene..2.7.dimethyl.	-0.10465	-0.16337
x22.26_Heptanal	0.162034	-0.01324
x22.52_1.6.Octadiene..3.7.dimethyl....S..	-0.10819	-0.16034
x22.97_Pentanoic.acid	-0.07487	-0.10194
x23.51_Methional	-0.01745	-0.06163
x23.84_Benzene..propyl.	-0.05209	-0.11159
x24.21_Decane	-0.068	-0.16648
x24.37_2.Heptanone..6.methyl.	0.127783	-0.04981
x24.41_beta.Pinene	-0.01639	-0.13013
x24.46_2.6.Octadiene..2.6.dimethyl.	-0.05604	-0.02457
x24.68_2.Heptanone..5.methyl.	-0.06023	-0.13398
x24.74_Furan..2.pentyl.	0.170222	-0.05473
x24.81_2.Heptenal...Z..	0.135183	-0.00703
x25.14_1.Hepten.3.one	0.00015	-0.00479
x25.36_3.Octanone	0.068713	-0.13536

x25.44_Benzene..1.2.4.trimethyl.	-0.02459	-0.13617
x25.441_Benzaldehyde	0.149093	-0.02558
x25.55_5.Hepten.2.one..6.methyl.	0.15314	-0.04951
x25.58_Benzene..1.2.3.trimethyl.	-0.07991	-0.07396
x25.66_2.Octanone	0.079119	-0.14357
x26.01_Octanal	0.184741	-0.00296
x26.20_D.Limonene	-0.06085	-0.14301
x26.53_Dimethyl.sulfone	0.0066	0.019641
x26.63_Decane..2.methyl.	-0.06854	-0.0395
x27.06_gamma.terpinene	0.090934	-0.01826
x27.15_Benzene..1.methyl.3.propyl.	-0.12333	-0.12937
x27.66_Cyclohexanone..2.2.6.trimethyl.	0.104485	-0.16351
x27.70_Undecane	-0.03886	-0.09776
x28.00_Phenol	-0.03813	-0.07124
x28.64_Benzeneacetaldehyde	0.052531	-0.08572
x29.16_2.Nonanone	0.146283	-0.07676
x29.51_Nonanal	0.169579	0.0227
x29.99_Undecane..2.methyl.	0.006704	0.009654
x30.37_2.3H..Furanone..5.ethyldihydro.	0.076369	-0.01527
x30.91_p.Cresol	-0.00895	-0.00873
x31.43_Undecane..2.6.dimethyl.	-0.06852	-0.09289
x31.64_1.Nonanol	0.032729	-0.03646
x32.41_2.Decanone	0.077121	-0.05719
x32.76_Decanal	0.148505	0.069727
x33.78_Phenol..4.ethyl.	-0.02184	-0.12839
x34.03_Tridecane	-0.0379	-0.18141
x34.15_1.Cyclohexene.1.carboxaldehyde..2.6.6.trimethyl.	0.127245	-0.06409
x35.46_2.Undecanone	0.065965	-0.13532
x35.70_Undecanal	0.127606	-0.02063
x35.85_Tridecane..2.methyl.	-0.05811	-0.10566
x36.01_Tridecane..3.methyl.	0.000706	-0.12103
x36.30_Dodecane..2.6.10.trimethyl.	0.051042	-0.16318
x36.89_Tetradecane	-0.01447	-0.18166
x37.19_2.Hexadecene..3.7.11.15.tetramethyl....R..R..R...E....	0.057814	-0.10052
x37.80_2.Undecenal	0.111181	0.063976
x38.57_Indole	-0.13752	-0.01082
x38.70_Heptadecane..2.6.10.14.tetramethyl.	-0.00766	-0.15838
x38.91_Tetradecane..3.methyl.	-0.04844	-0.13641
x39.85_Pentadecane	0.045218	-0.15526
x6.62_Ethanol	0.093071	-0.01714
x7.19_Propanal	0.12338	-0.10469
x7.35_Acetone	0.142767	-0.03903
x7.48_Dimethyl.sulfide	0.108742	-0.09406
x7.55_Isopropyl.Alcohol	0.041971	0.026241
x7.78_Carbon.disulfide	0.055775	0.044384

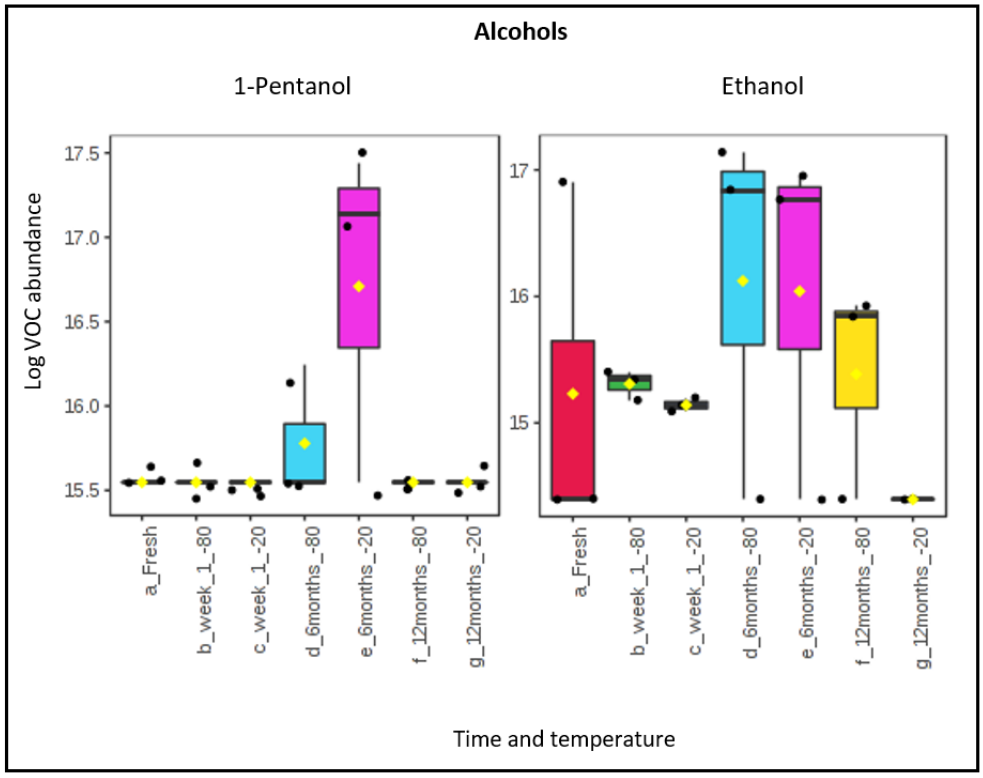
x7.92_Acetic.acid..methyl.ester	0.004836	-0.062
x8.98_Propanal..2.methyl.	0.125905	-0.07895
x9.54_1.Propanol	-0.00771	-0.11233
x9.89_Furan..2.methyl.	0.160797	-0.03517

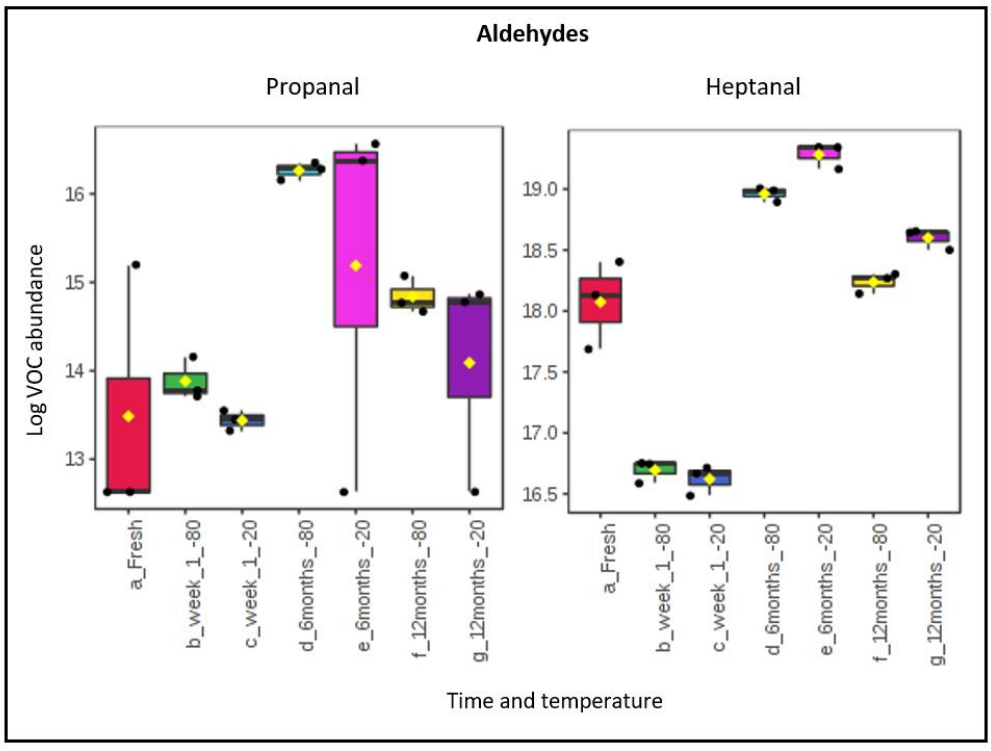
Appendix 2.6 Some key VOCs of various chemical classes were plotted to support Figure 2.12.



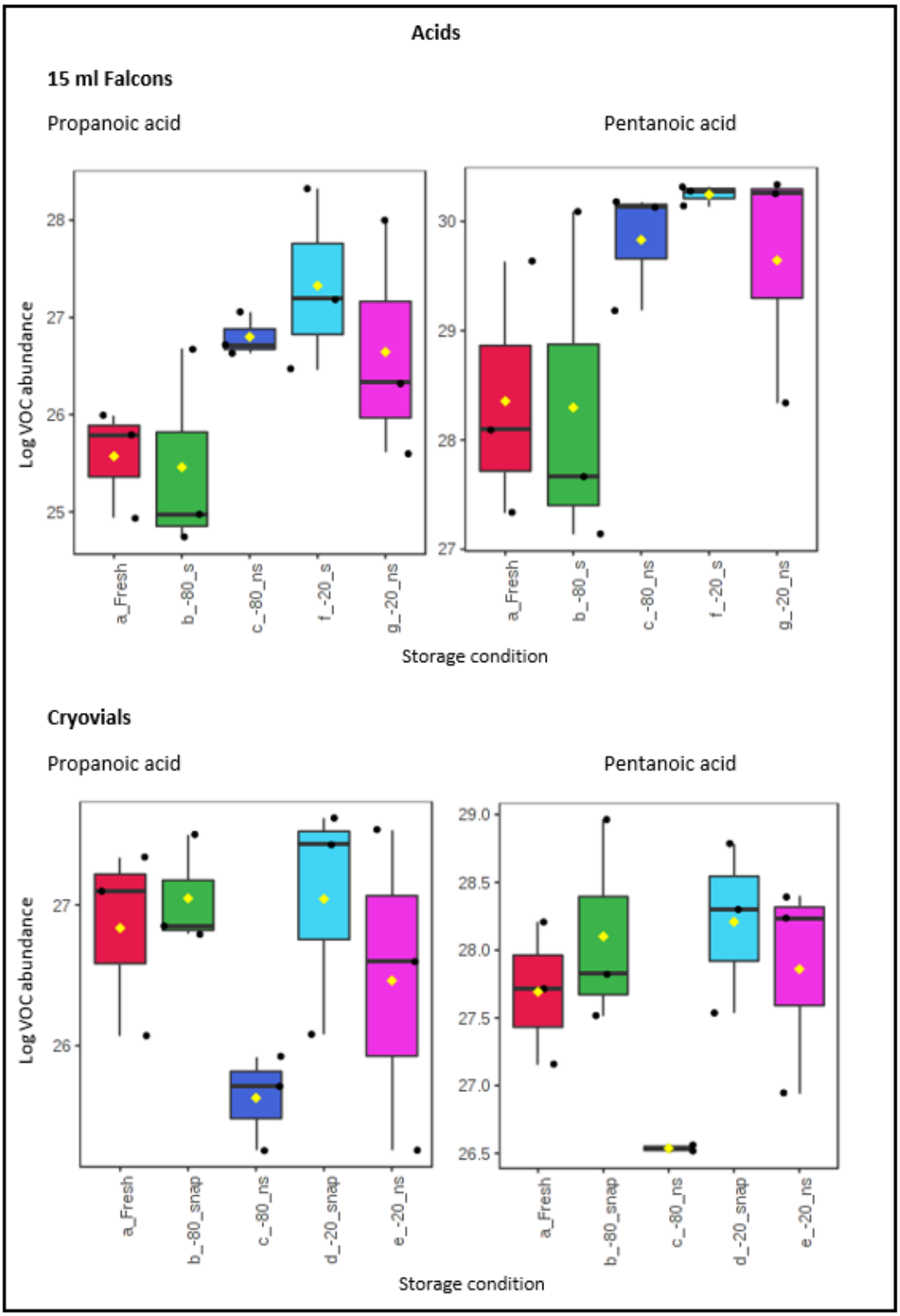


Appendix 2.6 Continued.

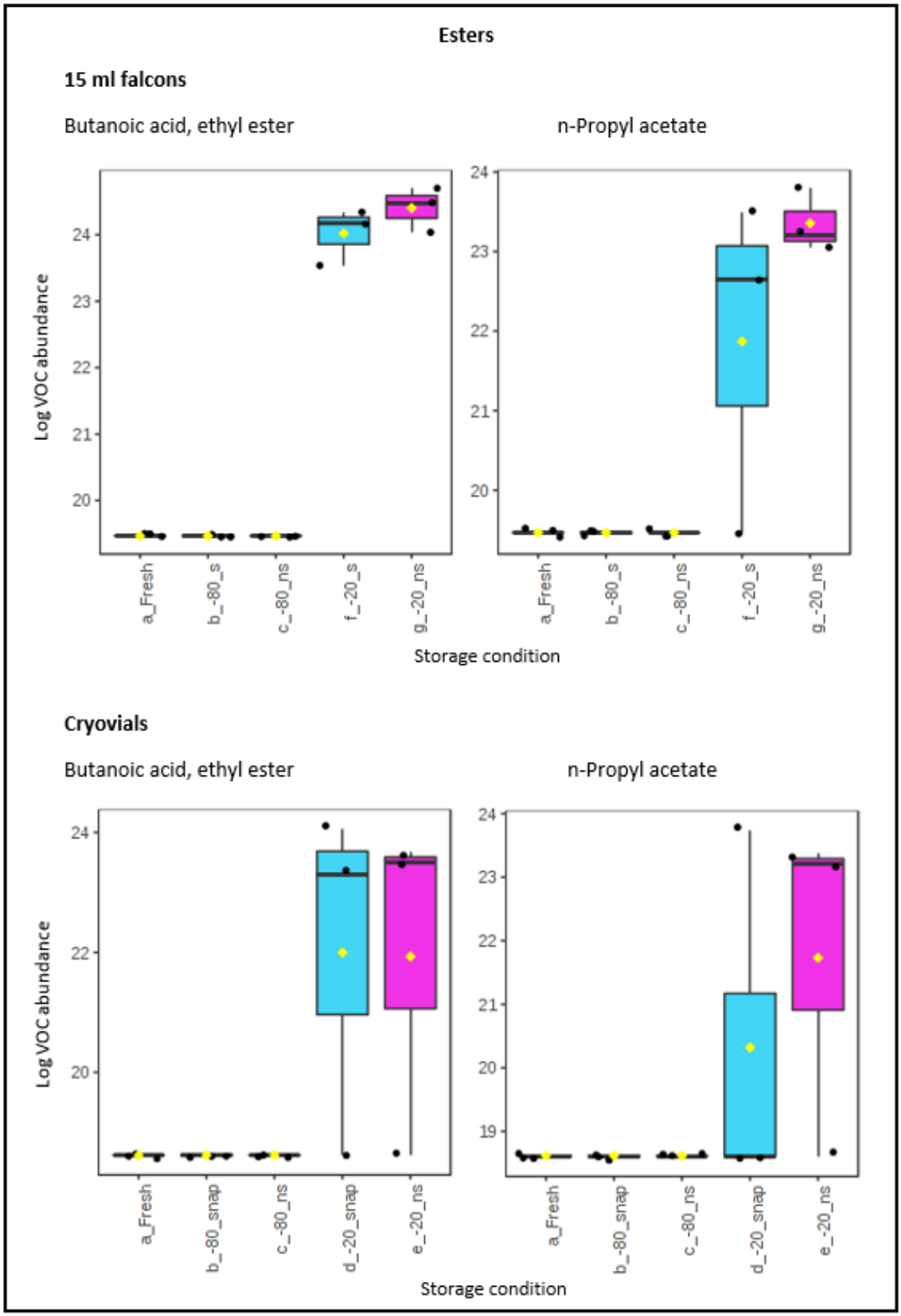




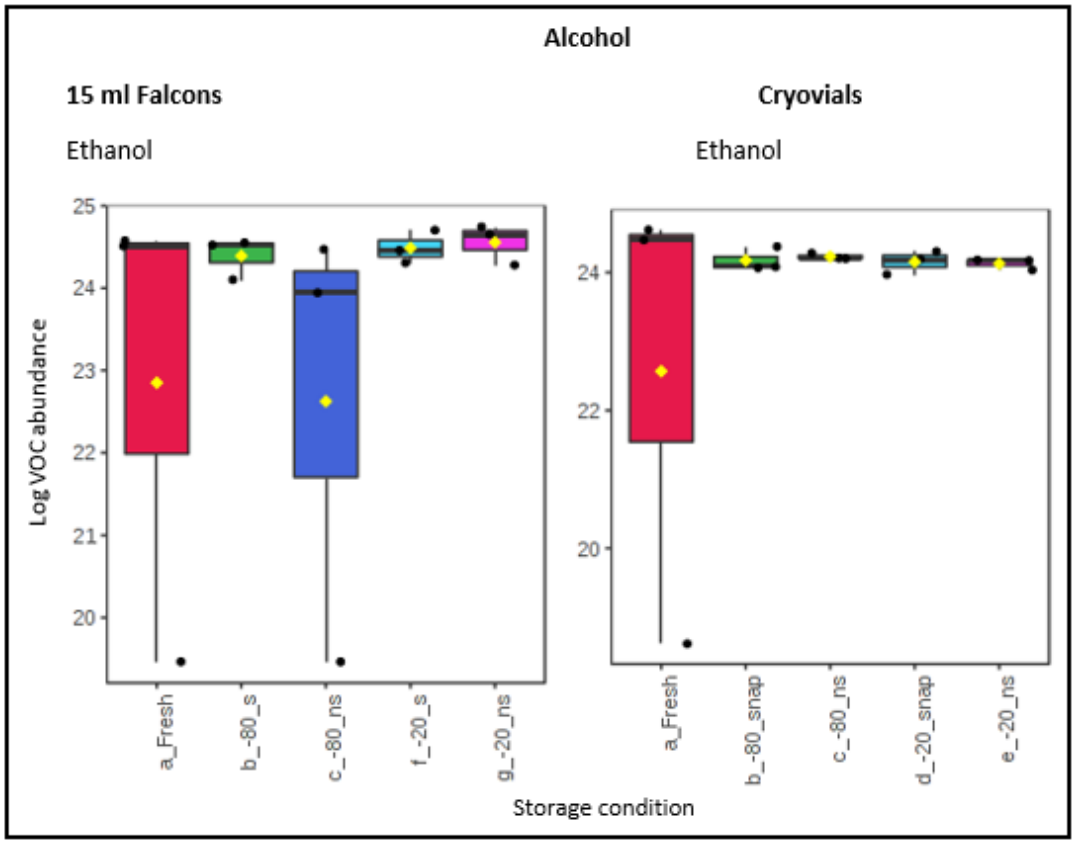
Appendix 2.7 Some key VOCs of various chemical classes were plotted to support Figure 2.12. Just one alcohol (ethanol) was shared between cryovials and 15 ml falcons. Key: s/snap = snap frozen in liquid nitrogen, ns = not snap frozen.



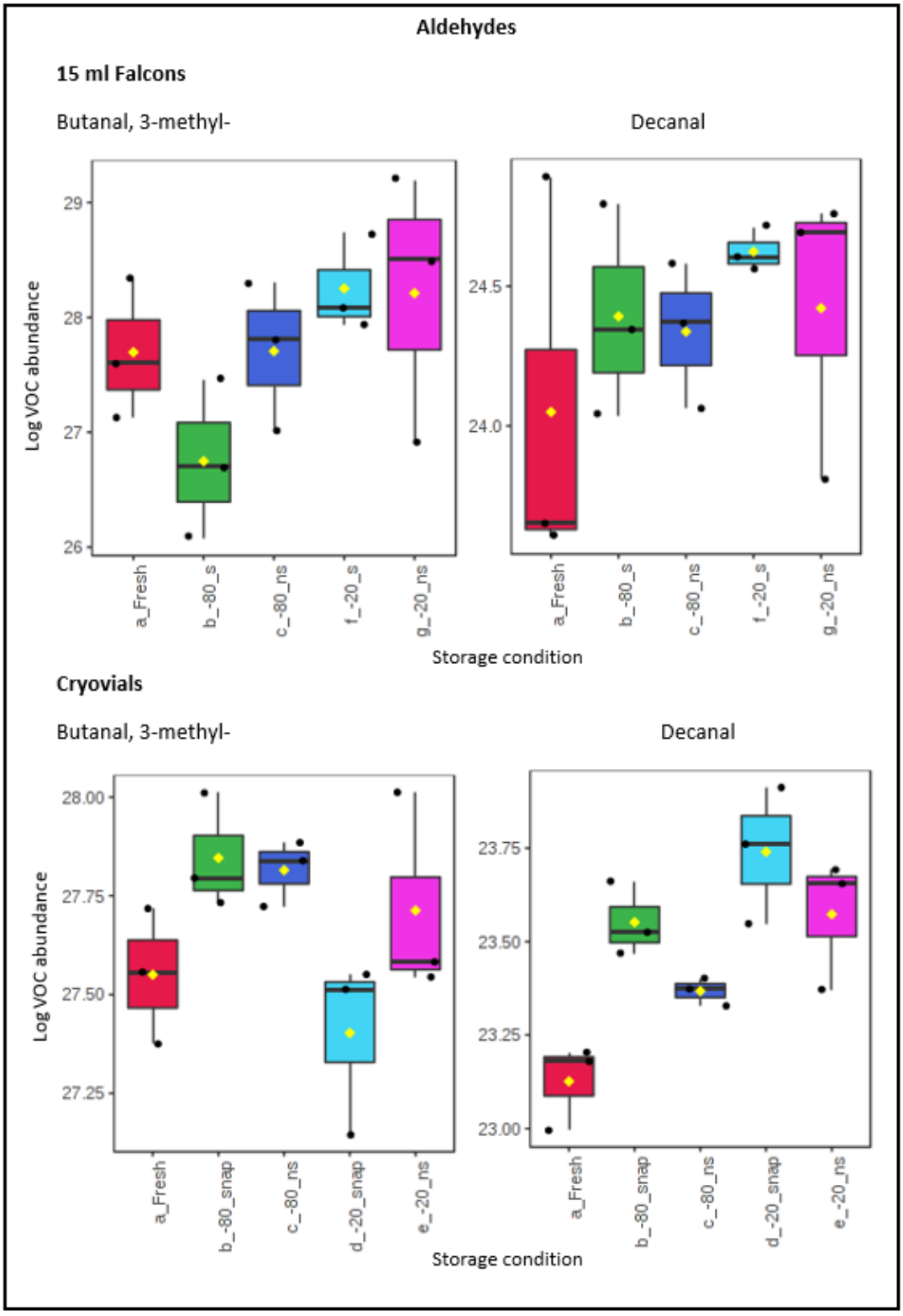
Appendix 2.7 Continued.



Appendix 2.7 Continued.



Appendix 2.7 Continued.



Appendix 3.1 The set of barcoded index primers.

Sample 18S	Primer set	Forward	Reverse
T15_1	N507_N706	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T15_2	N503_N708	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
T15_3	N508_N708	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
T15_4	N506_N707	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
T15_5	N502_N709	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
T15_7	N502_N708	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
T21_1	N508_N706	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T21_2	N508_N707	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
T21_3	N504_N708	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
T21_4	N505_N706	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T21_5	N507_N707	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
T21_7	N506_N708	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
T3_1	N504_N706	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T3_2	N506_N706	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T3_3	N507_N708	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
T3_4	N503_N709	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
T3_5	N503_N707	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
T3_7	N505_N708	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
T9_1	N505_N707	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
T9_2	N504_N707	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
T9_3	N502_N706	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T9_4	N503_N706	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC

T9_5	N502_N707	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
Sample ITS1	Primer set	Forward	Reverse
T15_1	N503_N710	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
T15_2	N506_N711	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTC
T15_3	N504_N712	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGACGTGTGCTC
T15_4	N502_N711	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTC
T15_5	N505_N712	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGACGTGTGCTC
T15_7	N505_N711	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTC
T21_1	N504_N710	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
T21_2	N504_N711	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTC
T21_3	N507_N711	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTC
T21_4	N508_N709	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
T21_5	N503_N711	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTC
T21_7	N502_N712	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGACGTGTGCTC
T3_1	N507_N709	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
T3_2	N502_N710	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
T3_3	N503_N712	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGACGTGTGCTC
T3_4	N506_N712	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTGACTGGAGTTCAGACGTGTGCTC
T3_5	N506_N710	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
T3_7	N508_N711	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTGACTGGAGTTCAGACGTGTGCTC
T9_1	N508_N710	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
T9_2	N507_N710	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
T9_3	N505_N709	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
T9_4	N506_N709	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
T9_5	N505_N710	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC

Appendix 3.2 List of core and inter-changeable compounds identified in faecal samples over 12 months

Retention time_VOC	April	May1	May2	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar1	Mar2
10.22_2,3-Butanedione	6	6	5	5	6	5	3	5	5	5	5	6	5	5
10.50_2-Butanone	4	2	2	0	1	4	4	5	5	4	3	2	2	6
10.60_Ethyl Acetate	6	6	6	6	4	1	3	2	2	2	5	4	6	0
11.15_Methyl propionate	4	6	6	6	4	4	6	4	5	3	2	4	5	1
11.93_1-Propanol, 2-methyl-	3	4	1	1	2	1	3	2	4	3	4	5	5	4
12.45_Butanal, 3-methyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
12.55_Acetic acid	6	6	4	6	5	3	3	5	4	4	5	5	4	2
12.79_Butanal, 2-methyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
13.05_Propanoic acid, 2-methyl-, methyl ester	0	1	1	2	0	3	1	0	1	1	0	0	0	0
13.30_1-butanol	1	5	2	2	1	0	0	1	2	3	1	0	1	1
13.43_Furan, 2-ethyl-	5	4	3	4	4	5	4	5	2	5	6	5	6	5
13.80_1-Penten-3-one	3	3	4	2	4	4	5	5	5	6	5	4	5	5
13.88_2-Pentanone	4	4	4	2	3	3	3	4	2	3	3	5	3	6
13.94_1-Penten-3-ol	1	3	2	1	2	4	4	5	5	3	0	3	3	1
14.10_Propanoic acid, ethyl ester	6	6	6	6	5	4	4	4	1	4	5	6	6	2
14.17_3-Pentanone	0	0	0	0	2	4	3	4	5	2	0	0	0	0
14.21_Pentanal	6	6	6	6	6	6	4	6	5	6	6	6	6	6
14.30_n-Propyl acetate	2	5	6	5	4	2	3	1	0	2	2	0	1	0
14.63_Butanoic acid, methyl ester	6	5	6	6	5	6	6	4	4	3	3	4	1	2
15.91_Propanoic acid, 2-methyl-, ethyl ester	3	4	2	6	4	3	1	2	1	4	5	5	2	1
16.07_Propanoic acid	6	5	6	6	6	6	5	5	6	6	4	5	4	4
16.23_1-Butanol, 2-methyl-	2	1	0	0	1	1	2	2	3	3	3	5	3	4
16.50_Toluene	6	6	6	6	6	6	6	6	6	6	6	6	6	6
16.66_Octane	6	3	0	0	0	1	1	6	4	4	6	6	6	3

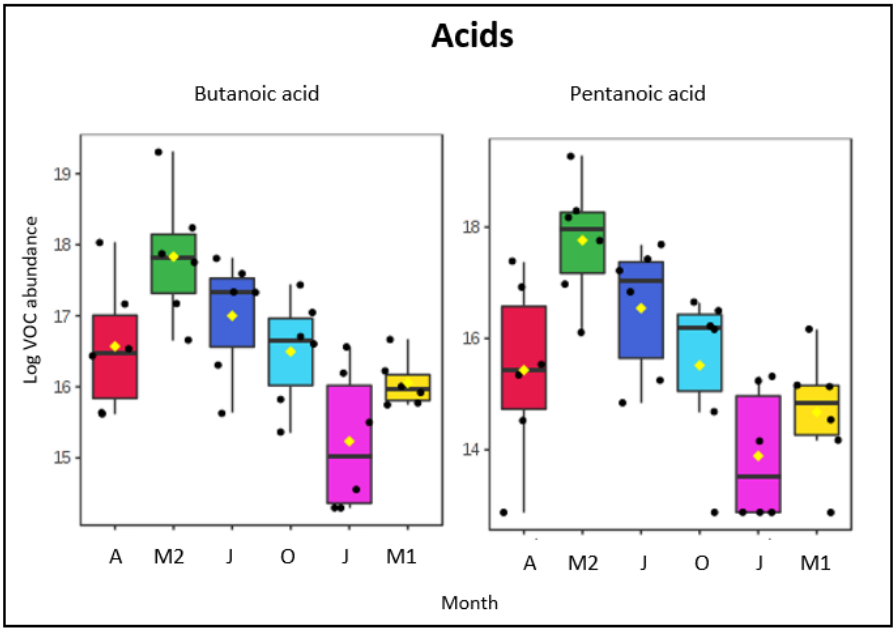
Retention time_VOC	April	May1	May2	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar1	Mar2
16.68_Butanoic acid, 2-methyl-, methyl ester	3	1	1	2	0	2	3	2	2	2	0	1	0	0
16.92_2-Octene, (E)-	6	6	6	6	6	6	6	6	6	5	5	6	5	5
16.93_1-Propanone, 1-cyclopropyl-	4	0	0	0	0	0	0	0	0	0	3	4	5	0
17.27_1-Pentanol	4	5	2	2	2	5	5	5	5	6	3	5	5	6
17.35_2-Octene	0	2	0	0	0	1	2	0	0	2	1	2	0	3
17.61_Butanoic acid, ethyl ester	6	6	6	5	6	6	6	6	5	6	5	6	5	5
17.99_Propanoic acid, propyl ester	5	6	6	6	5	5	5	1	3	4	4	4	4	1
18.01_2-Hexanone	2	1	1	2	1	1	1	5	3	0	0	1	2	4
18.25_Hexanal	6	6	6	6	6	6	6	6	6	6	6	6	6	6
18.36_Propanoic acid, 2-methyl-	1	3	3	4	3	2	3	0	3	1	1	0	0	1
18.64_Pentanoic acid, methyl ester	4	3	6	5	5	4	4	3	1	3	2	2	1	0
19.13_Butanoic acid, 1-methylethyl ester	2	6	6	6	5	4	5	1	5	4	0	0	1	0
19.40_Butanoic acid	6	6	6	6	6	6	6	6	6	6	6	6	6	5
19.45_Butanoic acid, 2-methyl-, ethyl ester	5	6	4	5	5	4	6	4	6	6	5	4	2	2
19.65_Propanoic acid, 2-methyl-, propyl ester	4	5	6	6	5	6	5	1	5	5	3	1	1	0
19.67_Butanoic acid, 3-methyl-, ethyl ester	5	6	6	6	6	6	6	6	6	6	5	6	5	3
20.17_Propanoic acid, 2-methylpropyl ester	5	6	6	6	5	5	4	2	4	3	3	6	2	2
20.30_Ethylbenzene	6	5	6	6	5	5	6	6	6	6	6	6	6	6
20.49_Nonane	6	6	6	6	6	6	6	6	6	6	6	6	6	6
20.60_Benzene, 1,3-dimethyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
20.83_1-Butanol, 3-methyl-, acetate	1	5	6	6	5	3	4	1	2	1	2	3	4	1
20.94_1-Butanol, 2-methyl-, acetate	2	4	6	5	3	3	3	1	2	2	2	1	0	0
21.00_2-Hexenal	6	6	6	6	6	6	6	5	6	6	6	6	6	6
21.32_1-Hexanol	4	6	2	1	2	5	5	5	5	6	3	5	5	6
21.44_Butanoic acid, 3-methyl-	6	6	4	4	6	5	6	6	6	5	5	3	4	5
21.69_Propanoic acid, butyl ester	0	6	6	5	5	2	2	1	2	3	1	0	1	0
21.71_Styrene	6	6	5	6	6	6	6	6	6	6	6	6	6	6
21.73_Butanoic acid, 2-methyl-	4	5	5	5	5	4	5	5	5	5	4	5	5	3

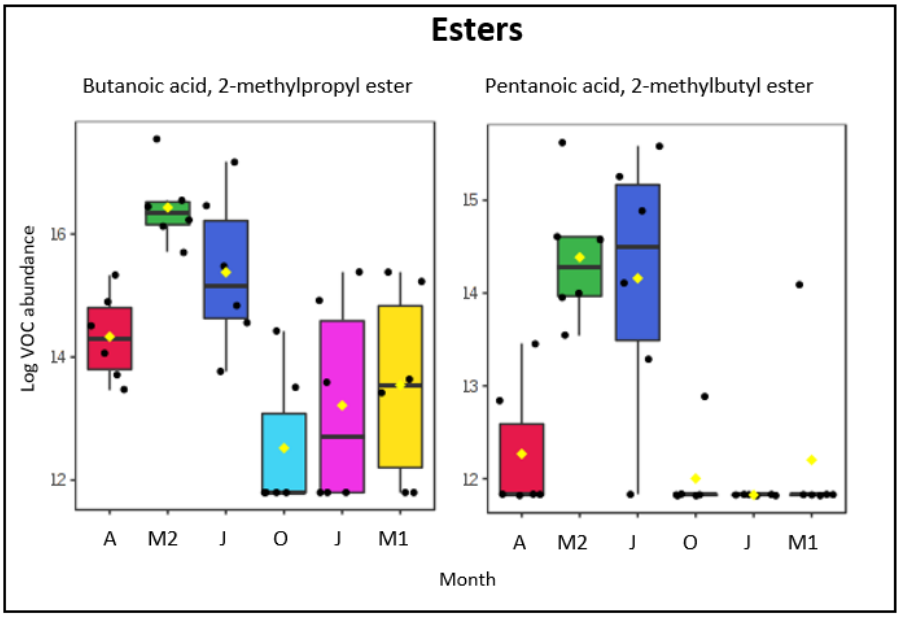
Retention time_VOC	April	May1	May2	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar1	Mar2
21.94_2-Heptanone	6	5	6	5	6	6	6	6	6	6	6	6	6	6
22.05_1,7-Octadiene, 2,7-dimethyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
22.26_Heptanal	6	6	6	6	6	6	6	6	6	6	6	6	6	6
22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
22.73_Pentanoic acid, 1-methylethyl ester	0	1	3	6	4	1	3	0	0	1	0	0	0	0
22.85_Nonane, 2-methyl-	6	4	0	4	3	5	5	6	6	3	6	6	4	6
22.97_Pentanoic acid	5	4	6	6	6	5	6	5	5	5	5	5	5	5
23.22_Butanoic acid, 3-methyl-, propyl ester	0	2	1	1	3	0	1	0	0	1	0	0	0	0
23.27_2-Octene, 2,6-dimethyl-	4	6	6	6	6	6	6	6	6	4	2	0	2	6
23.51_Methional	4	4	5	6	4	4	4	3	6	5	1	1	1	3
23.58_Butanoic acid, 2-methylpropyl ester	6	6	6	5	4	4	4	2	3	3	4	6	4	0
23.84_Benzene, propyl-	2	1	0	0	3	0	0	3	1	0	2	2	1	2
24.06_Oxime-, methoxy-phenyl-_	1	0	0	0	0	1	1	1	1	4	3	3	4	1
24.12_cis-2,6-Dimethyl-2,6-octadiene	2	4	4	5	4	6	5	6	4	3	1	2	0	5
24.21_Decane	5	6	3	5	5	5	5	6	6	6	6	6	6	6
24.37_2-Heptanone, 6-methyl-	6	6	6	6	6	5	6	6	6	6	6	6	6	6
24.41_beta-Pinene	2	0	1	0	1	3	2	4	0	2	5	0	3	0
24.46_2,6-Octadiene, 2,6-dimethyl-	4	6	4	4	5	2	6	5	3	3	4	3	2	4
24.68_2-Heptanone, 5-methyl-	6	5	4	6	5	6	4	6	6	6	6	6	6	6
24.74_Furan, 2-pentyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
24.81_2-Heptenal, (Z)-	2	2	1	1	1	1	1	1	1	3	3	1	1	3
24.90_1-Heptanol	4	3	0	0	1	3	2	3	5	4	4	5	5	6
24.93_Pentanoic acid, propyl ester	1	1	4	3	2	1	3	1	0	0	0	0	1	0
25.14_1-Hepten-3-one	1	0	0	3	3	4	3	5	6	4	1	0	0	6
25.23_1-Octen-3-ol	4	6	5	6	5	6	6	6	6	6	3	5	5	6
25.33_1-Butanol, 3-methyl-, propanoate	4	5	6	3	3	1	2	0	1	0	4	6	2	0
25.36_3-Octanone	6	6	5	6	6	6	6	6	6	6	6	6	6	6
25.44_Benzene, 1,2,4-trimethyl-	1	2	0	1	2	0	2	4	3	2	2	2	2	3

Retention time_VOC	April	May1	May2	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar1	Mar2
25.441_Benzaldehyde	6	6	6	6	6	6	6	6	6	6	6	6	6	6
25.55_5-Hepten-2-one, 6-methyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
25.66_2-Octanone	1	0	0	3	2	0	1	2	0	0	2	1	1	1
25.68_3-Octanol	4	6	5	4	3	6	6	4	5	6	3	5	4	6
26.01_Octanal	6	6	6	6	6	6	6	6	6	6	6	6	6	6
26.20_D-Limonene	6	6	6	6	6	6	6	6	6	5	6	6	6	6
26.53_Dimethyl sulfone	5	3	2	3	2	1	1	0	5	3	4	4	5	4
27.06_gamma terpinene	3	4	1	2	4	6	1	3	2	1	1	2	1	1
27.14_Butanoic acid, 2-methylbutyl ester	6	6	6	6	4	5	4	3	4	4	4	6	5	3
27.66_Cyclohexanone, 2,2,6-trimethyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
27.70_Undecane	6	6	6	6	6	6	6	6	6	6	6	6	6	6
28.00_Phenol	6	6	6	6	6	6	6	6	6	6	5	5	5	6
28.27_Pentanoic acid, butyl ester	3	5	6	6	6	4	6	1	5	6	0	0	2	0
28.64_Benzeneacetaldehyde	6	6	6	6	6	6	6	6	6	6	6	6	6	6
29.00_Propanoic acid, hexyl ester	2	3	2	0	0	0	0	0	0	0	0	1	0	0
29.16_2-Nonanone	5	5	6	6	6	6	6	6	6	4	5	5	4	5
29.51_Nonanal	6	6	6	6	6	6	6	6	6	6	6	6	6	6
29.81_Decane, 2,9-dimethyl-	0	1	3	4	1	2	1	0	3	0	3	3	1	5
29.99_Undecane, 2-methyl-	4	5	3	4	6	3	5	5	3	6	4	5	5	2
30.32_Pentanoic acid, 3-methylbutyl ester	4	5	6	5	4	3	5	3	3	3	1	3	2	0
30.37_2(3H)-Furanone, 5-ethyldihydro-	6	5	5	6	6	5	5	6	6	6	6	6	6	6
30.44_Pentanoic acid, 2-methylbutyl ester	1	3	6	5	4	3	4	0	1	1	0	1	1	0
30.91_p-Cresol	6	6	6	6	6	6	6	6	6	6	6	6	6	6
31.21_Phenylethyl Alcohol	2	4	5	2	1	4	4	3	5	5	1	4	3	2
31.43_Undecane, 2,6-dimethyl-	5	5	5	6	6	6	6	6	6	6	6	6	6	6
31.64_1-Nonanol	4	3	6	2	1	3	2	5	4	4	3	5	6	4
32.41_2-Decanone	6	6	6	6	6	6	6	6	6	5	6	6	6	5
32.76_Decanal	6	6	6	6	6	6	6	6	6	6	6	6	6	6

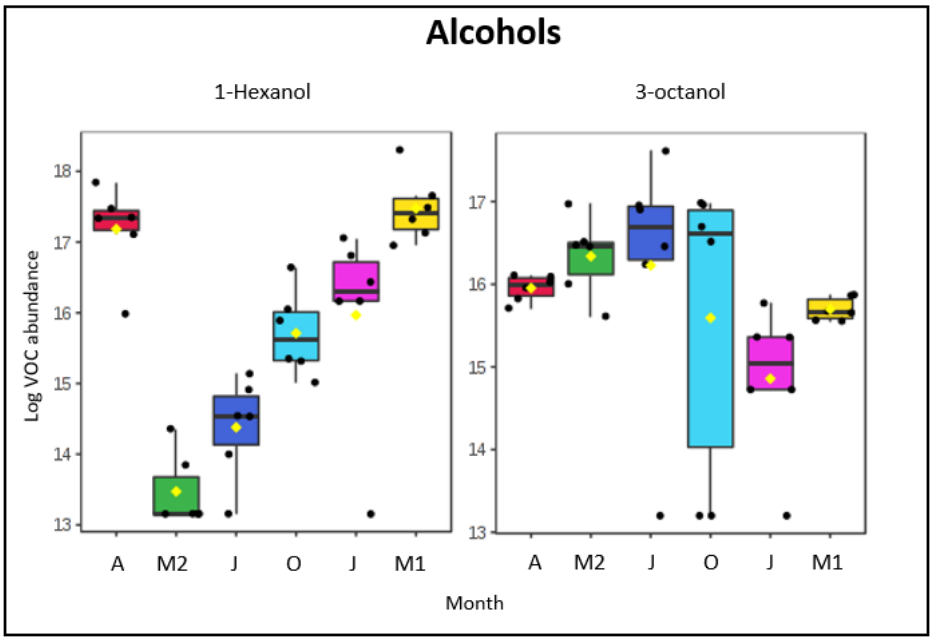
Retention time_VOC	April	May1	May2	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar1	Mar2
32.86_Dodecane, 2-methyl-	5	5	5	6	6	6	6	6	5	6	6	6	6	6
33.78_Phenol, 4-ethyl-	5	6	6	6	6	6	6	6	6	6	6	5	6	6
34.03_Tridecane	6	6	6	6	6	6	6	6	6	6	6	6	6	6
34.15_1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
35.46_2-Undecanone	6	6	6	6	6	6	6	6	6	6	6	6	6	6
35.70_Undecanal	5	2	0	1	0	0	0	0	5	4	4	3	3	0
35.73_1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	2	3	6	6	5	1	6	1	5	4	0	0	1	1
35.85_Tridecane, 2-methyl-	6	5	6	6	6	6	6	6	6	5	6	6	6	6
36.01_Tridecane, 3-methyl-	6	5	6	6	6	6	6	6	6	6	6	6	6	6
36.30_Dodecane, 2,6,10-trimethyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
36.34_3-Buten-2-one, 4-(2-hydroxy-2,6,6-trimethylcyclohexyl)-	0	1	0	0	3	1	0	0	1	0	0	0	0	0
36.36_Phenol, 4-propyl-	1	1	0	0	0	0	0	0	0	0	0	2	1	1
36.89_Tetradecane	6	6	6	6	6	6	6	6	6	6	6	6	6	6
37.19_2-Hexadecene, 3,7,11,15-tetramethyl-, [R-[R*,R*-(E)]]-	5	3	4	5	3	3	2	4	4	4	3	5	3	4
37.34_unknown_compound	0	0	0	3	0	0	0	0	0	0	0	0	0	0
37.80_2-Undecenal	3	5	6	2	2	4	2	4	6	4	1	2	2	0
38.57_Indole	3	5	6	5	5	3	4	1	3	5	3	1	5	5
38.70_Heptadecane, 2,6,10,14-tetramethyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
38.91_Tetradecane, 3-methyl-	5	5	4	4	6	6	5	4	4	3	6	6	6	6
39.85_Pentadecane	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7.19_Propanal	1	1	0	0	1	0	0	0	1	1	0	2	3	2
7.35_Acetone	1	1	0	1	1	3	1	1	1	2	0	1	1	2
7.48_Dimethyl sulfide	2	3	2	1	1	0	1	0	2	2	0	0	3	3
7.92_Acetic acid, methyl ester	0	1	2	0	3	1	0	0	0	1	0	0	1	1
8.98_Propanal, 2-methyl-	6	6	6	6	6	6	6	6	6	6	6	6	6	6
9.89_Furan, 2-methyl-	1	1	0	0	3	2	4	5	1	0	0	1	1	2

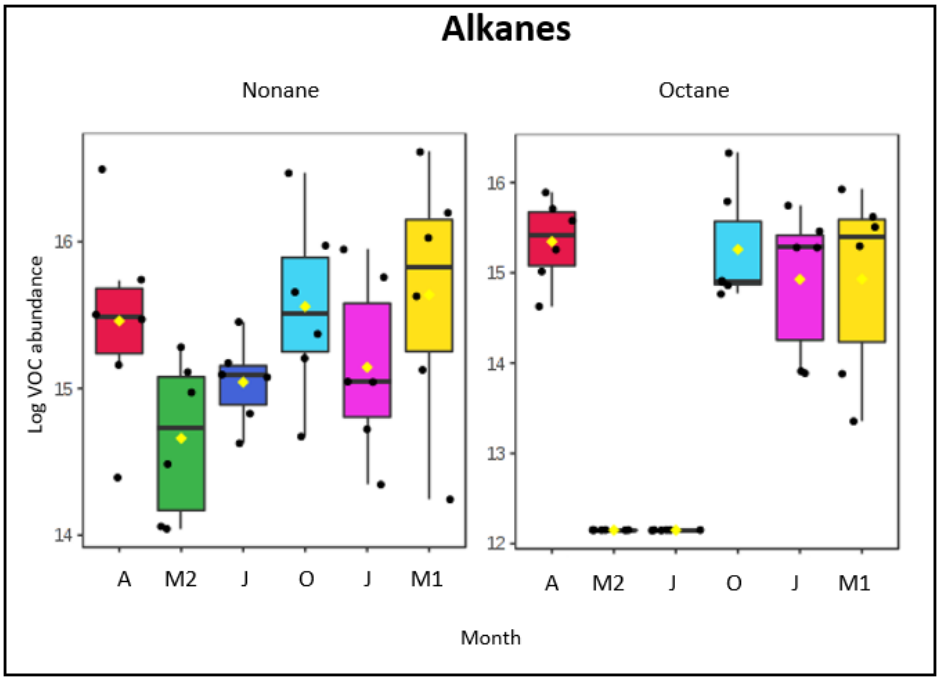
Appendix 3.3 Boxplots of some key VOCs of various chemical classes to support Figure 3.3.



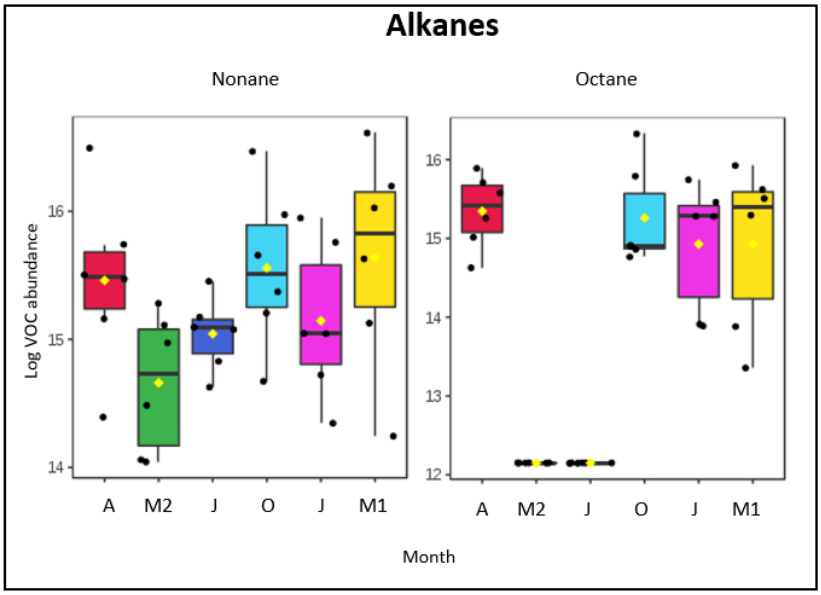


Appendix 3.3 Continued.

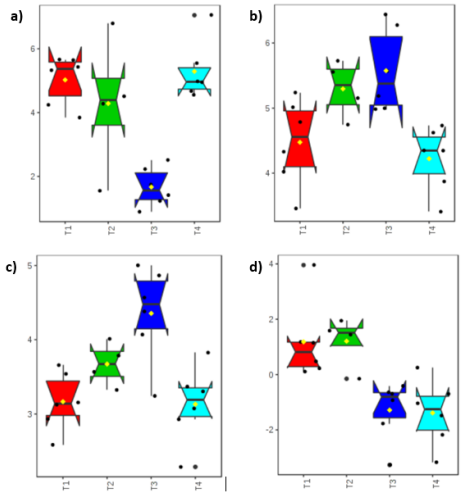




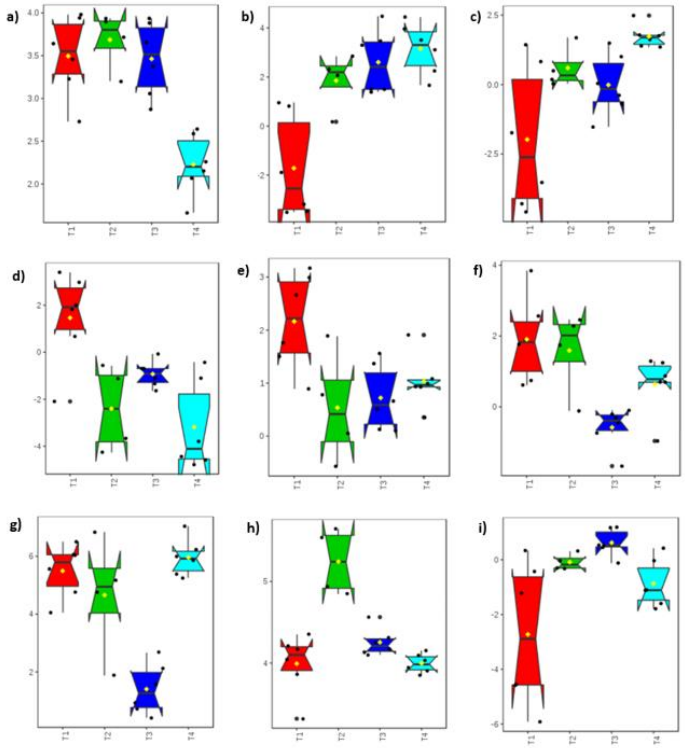
Appendix 3.3 Continued.



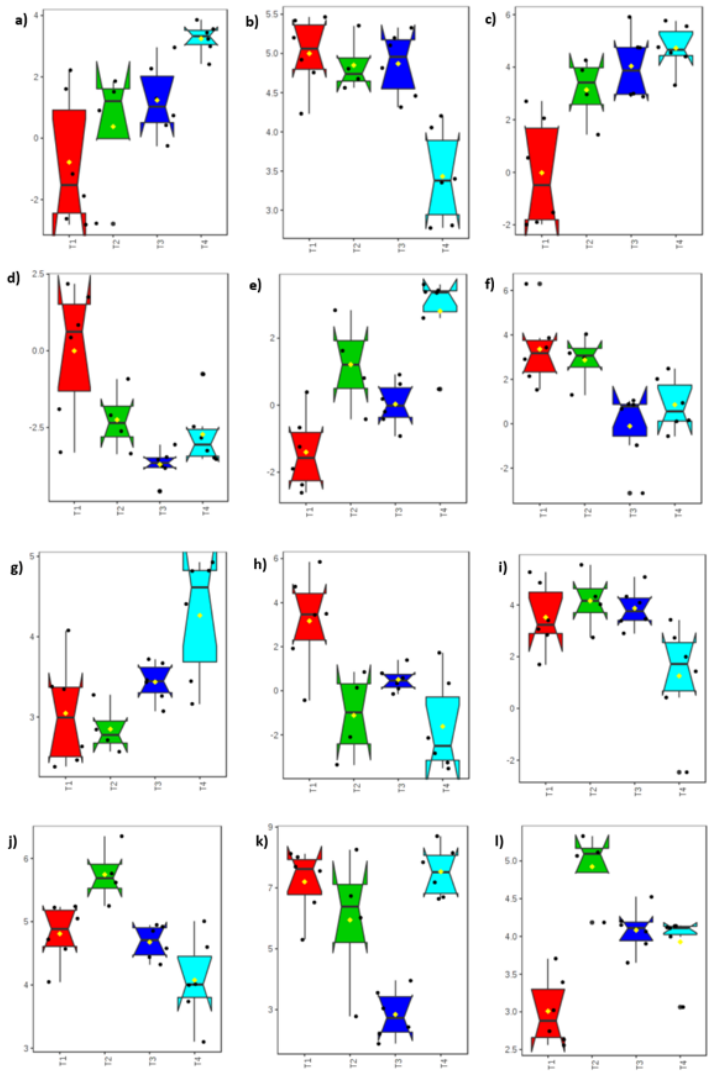
Appendix 3.4a Boxplots of Phylum associated with feed type and time (18S rRNA). a= neocallimastigomycota. b= Ascomycota, c= basidomycota, d= zygomycota



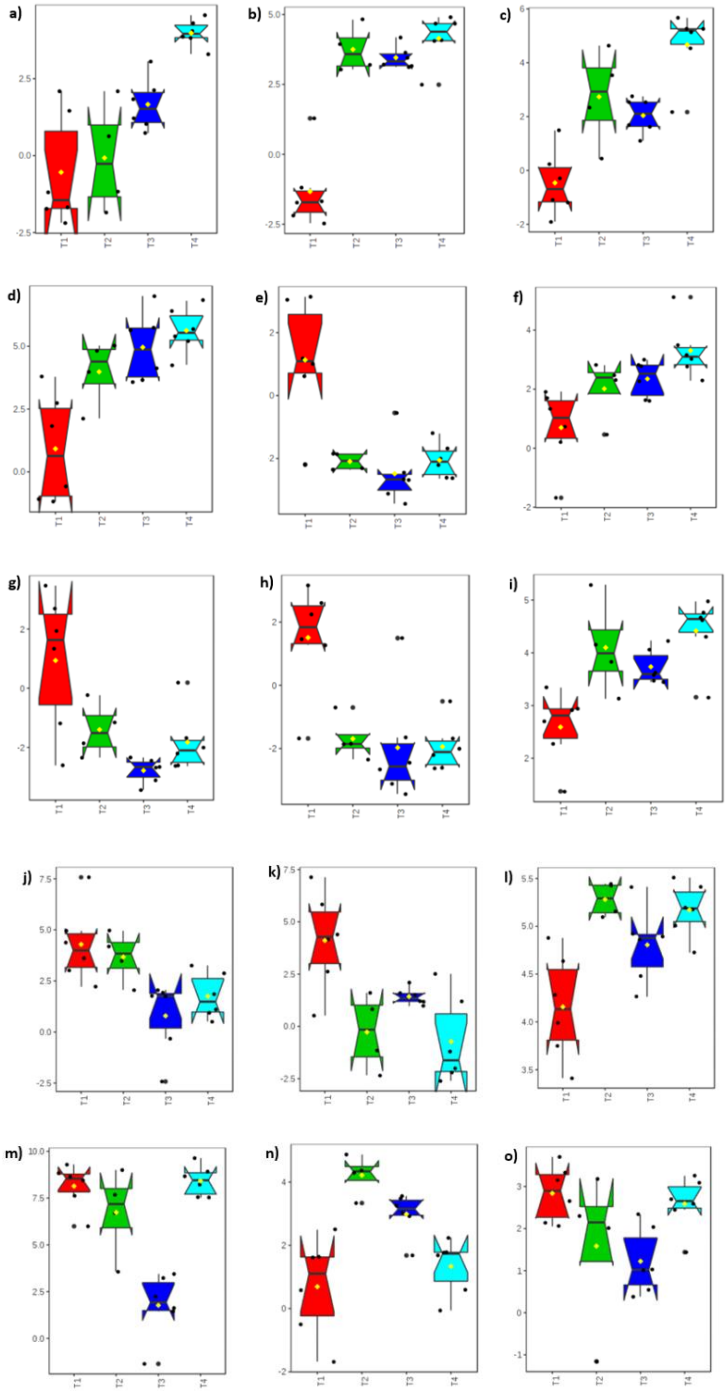
Appendix 3.4b Boxplots of Class associated with feed type and time (18S rRNA). a= Tremellomycetes, b= Pucciniomycetes, c= Eurotiomycetes, d= Taphrinomycetes, e=Pezizomycetes, f= Mucoromycotina, g= Neocallimastigomycetes, h= Dothideomycetes, i= Cystobasidiomycetes



Appendix 3.4c Boxplots of Order associated with feed type and time (18S rRNA). a= Chaetothyriales, b= Tremellales, c= Pucciniales, d= Microbotryales, e= Glomerellales, f= Mucorales, g= Helotiales, h= Taphrinales, i= Sordariales , j= Capnodiales, k= Hypocreales



Appendix 3.4d Boxplots of Family associated with feed type and time (18S rRNA). a= Chaetothyriaceae, b= Plectosphaerellaceae, c= Glomerellaceae, d= Pucciniaceae, e= Erysiphaceae, f= Lyophyllaceae, g= Ustilentylomataceae, h= Debaryomycetaceae, i= Hypocreales Incertae Sedis, j= Pilobolaceae, k= Protomycetaceae, l= Leptosphaeriaceae, m= Neocallimastigaceae, n= Coniothyriaceae



Appendix 3.5 Labels of OTUs in loadings for Pearson's correlation plot and circos plot (18S rRNA)

	Phylum	Class	Order	Family	Genus
OTU4	Ascomycota	Leotiomycetes	Helotiales		
OTU5	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	Fusarium
OTU6	Ascomycota	Leotiomycetes	Thelebolales		
OTU7	Ascomycota	Pezizomycetes	Pezizales	Ascobolaceae	
OTU8	Basidiomycota	Pucciniomycetes	Pucciniales	Pucciniaceae	Puccinia
OTU9	Ascomycota	Pezizomycetes	Pezizales	Pyronemataceae	Lasiobolidium
OTU10	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	
OTU12	Mucoromycota	Mucoromycotina	Mortierellales	Mortierellaceae	Mortierella
OTU13	Ascomycota	Saccharomycetes	Saccharomycetales		
OTU14	Zygomycota	Mucoromycotina	Mucorales	Pilobolaceae	Pilobolus
OTU16	Ascomycota	Sordariomycetes	Sordariomycetes Incertae Sedis	Plectosphaerellaceae	Plectosphaerella
OTU19	Basidiomycota	Agaricomycetes			
OTU20	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Guehomyces
OTU21	Zygomycota	Mucoromycotina	Mucorales	Pilobolaceae	Pilaira
OTU22	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	Penicillium
OTU23	Basidiomycota	Tremellomycetes	Tremellales	Tremellaceae	Cryptococcus
OTU25	Ascomycota	Eurotiomycetes	Chaetothyriales	Chaetothyriaceae	Cyphellophora
OTU26	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Sarocladium
OTU27	Ascomycota	Leotiomycetes	Helotiales		
OTU28	Blastocladiomycota	Blastocladiomycetes	Blastocladiales	Physodermataceae	Urophlyctis
OTU29	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetaceae	Kazachstania
OTU31	Basidiomycota	Tremellomycetes	Tremellales		
OTU32	Ascomycota	Sordariomycetes	Diaporthales	Valsaceae	Valsa
OTU33	Ascomycota	Sordariomycetes	Sordariales		
OTU34	Ascomycota	Dothideomycetes	Pleosporales		
OTU37	Basidiomycota	Agaricomycetes	Agaricales	Omphalotaceae	Lentinula
OTU38	Basidiomycota	Tremellomycetes	Tremellales		
OTU40	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales		
OTU41	Basidiomycota	Agaricomycetes			
OTU42	Ascomycota	Taphrinomycetes	Taphrinales	Protomycetaceae	Protomyces
OTU43	Basidiomycota	Microbotryomycetes	Microbotryales	Ustilentylomataceae	Ustilentyloma
OTU44	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Acremonium
OTU45	Basidiomycota	Cystobasidiomycetes	Cystobasidiales	Cystobasidiaceae	Cystobasidium
OTU47	Basidiomycota	Exobasidiomycetes	Exobasidiomycetes Incertae Sedis	Exobasidiomycetes Incertae Sedis	Tilletiopsis
OTU48	Ascomycota	Sordariomycetes	Calosphaeriales	Pleurostomataceae	Pleurostomophora
OTU49	Basidiomycota	Tremellomycetes	Tremellales		
OTU50	Ascomycota	Saccharomycetes	Saccharomycetales	Wickerhamomyceteae	Wickerhamomyces
OTU51	Ascomycota	Sordariomycetes	Sordariales	Sordariales incertae sedis	Kionochaeta
OTU53	Unclassified				
OTU55	Ascomycota	Leotiomycetes	Helotiales	Dermateaceae	Marssonia

OTU56	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	
OTU58	Ascomycota	Sordariomycetes			
OTU61	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Cystofilobasidium
OTU62	Basidiomycota	Agaricostilbomycetes	Agaricostilbales	Chionosphaeraceae	
OTU63	Ascomycota	Sordariomycetes	Sordariales		
OTU64	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Acremonium
OTU65	Ascomycota	Leotiomycetes	Erysiphales	Erysiphaceae	Blumeria
OTU68	Basidiomycota	Exobasidiomycetes	Entylomatales	Entylomataceae	Entyloma
OTU69	Unclassified				
OTU70	Chytridiomycota	Chytridiomycetes	Chytridiales		
OTU71	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	Fusarium
OTU75	Basidiomycota	Tremellomycetes	Tremellales		
OTU76	Ascomycota	Leotiomycetes	Helotiales	Helotiales Incertae Sedis	Pilidium
OTU77	Chytridiomycota				
OTU78	Ascomycota	Sordariomycetes	Ophiostomatales	Sordariomycetidae	Pesotum
OTU86	Chytridiomycota				
OTU87	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Ramularia
OTU90	Ascomycota	Pezizomycetes	Pezizales	Ascobolaceae	Ascobolus
OTU91	Ascomycota	Dothideomycetes	Pleosporales	Venturiaceae	Venturia
OTU92	Chytridiomycota				
OTU98	Basidiomycota	Agaricomycetes			
OTU99	Ascomycota	Dothideomycetes	Pleosporales		
OTU105	Ascomycota	Leotiomycetes	Helotiales		
OTU108	Ascomycota	Dothideomycetes	Pleosporales		
OTU111	Ascomycota	Pezizomycetes	Pezizales	Ascobolaceae	Thecotheus
OTU114	Ascomycota	Sordariomycetes	Microascales	Microascaceae	Scedosporium
OTU115	Ascomycota	Dothideomycetes	Acrospermales	Acrospermaceae	Acrospermum
OTU118	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Udeniomyces
OTU119	Ascomycota	Saccharomycetes	Saccharomycetales	Dipodascaceae	
OTU120	Ascomycota	Sordariomycetes	Sordariomycetes Incertae Sedis	Plectosphaerellaceae	Lectera
OTU123	Ascomycota	Eurotiomycetes	Chaetothyriales		
OTU127	Zygomycota	Mucoromycotina	Mucorales	Pilobolaceae	Pilobolus
OTU128	Ascomycota	Dothideomycetes	Myriangiales	Elsinoaceae	Elsinoe
OTU133	Basidiomycota	Tremellomycetes	Tremellales	Tremellaceae	Cryptococcus
OTU134	Ascomycota	Sordariomycetes	Xylariales	Amphisphaeriaceae	Pestalotiopsis
OTU135	Ascomycota	Pezizomycetes	Pezizales	Ascobolaceae	
OTU137	Ascomycota	Pezizomycetes	Pezizales	Ascobolaceae	
OTU138	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetales Incertae Sedis	Candida
OTU139	Ascomycota	Dothideomycetes	Dothideomycetes Incertae Sedis	Dothideomycetes Incertae Sedis	Peltaster
OTU140	Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	Aureobasidium
OTU141	Ascomycota	Sordariomycetes	Sordariales		
OTU147	Ascomycota	Leotiomycetes	Thelebolales		
OTU150	Ascomycota	Eurotiomycetes	Onygenales	Gymnoascaceae	Arachniotus

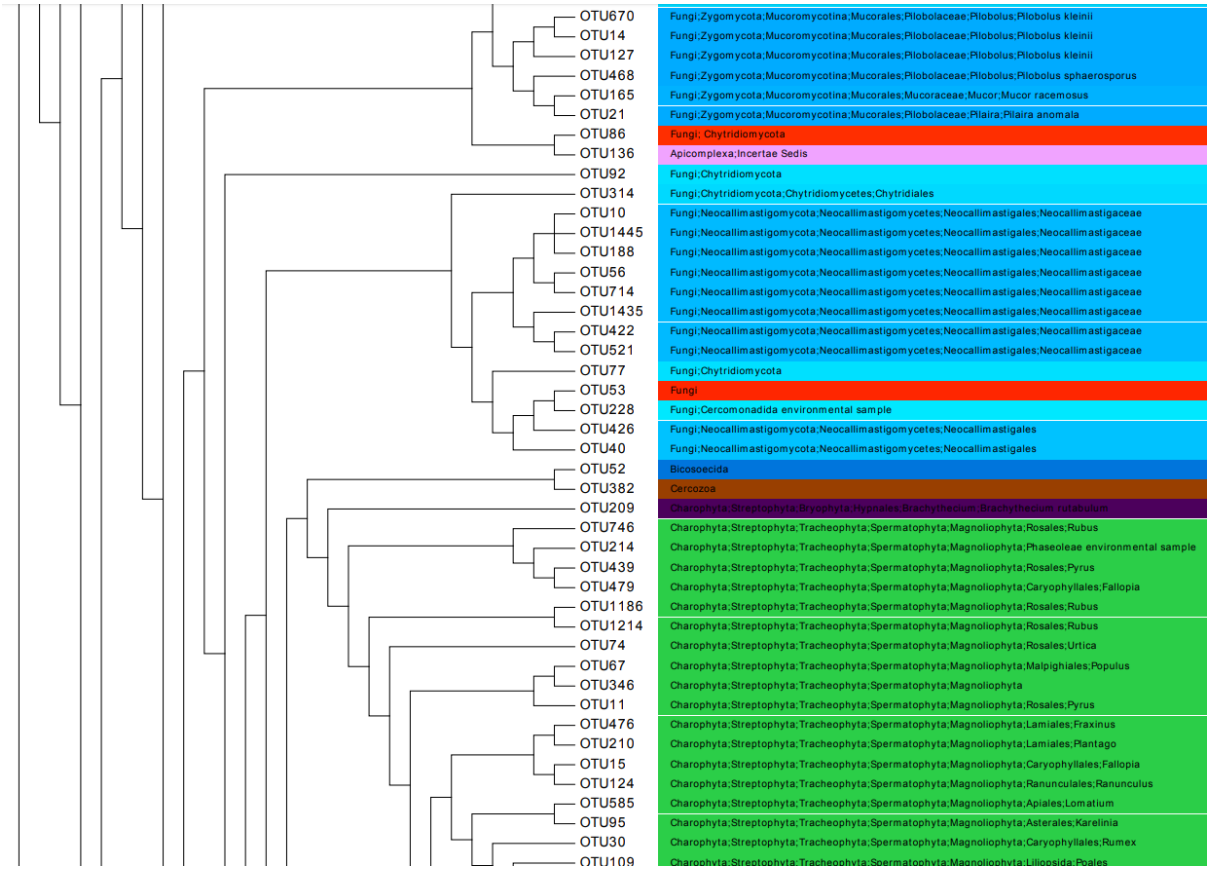
OTU152	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Acremonium
OTU154	Ascomycota	Eurotiomycetes	Pyrenulales	Rhynchostomataceae	Rhynchostoma
OTU156	Basidiomycota	Tremellomycetes	Tremellales		
OTU157	Basidiomycota	Cystobasidiomycetes	Cystobasidiales	Cystobasidiaceae	Occultifur
OTU164	Basidiomycota	Microbotryomycetes	Sporidiobolales	Sporidiobolaceae	Sporobolomyces
OTU165	Zygomycota	Mucoromycotina	Mucorales	Mucoraceae	Mucor
OTU166	Ascomycota	Saccharomycetes	Saccharomycetales		
OTU167	Ascomycota	Leotiomycetes	Helotiales		
OTU173	Ascomycota	Arthoniomycetes			
OTU176	Ascomycota	Dothideomycetes	Pleosporales		
OTU177	Basidiomycota	Agaricostilbomycetes	Agaricostilbales	Chionosphaeraceae	Kurtzmanomyces
OTU179	Basidiomycota	Agaricostilbomycetes	Agaricostilbales	Agaricostilbaceae	Bensingtonia
OTU180	Ascomycota	Leotiomycetes	Helotiales		
OTU185	Ascomycota	Dothideomycetes	Pleosporales	Melanommataceae	Bimuria
OTU188	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	
OTU189	Ascomycota	Sordariomycetes	Sordariomycetes Incertae Sedis	Plectosphaerellaceae	Verticillium
OTU190	Basidiomycota	Agaricomycetes	Auriculariales	Auriculariaceae	Auricularia
OTU191	Ascomycota	Dothideomycetes	Capnodiales		
OTU192	Basidiomycota	Pucciniomycetes	Pucciniales	Pucciniaceae	Puccinia
OTU194	Glomeromycota	Glomeromycetes	Glomerales	Glomeraceae	Rhizophagus
OTU197	Ascomycota	Sordariomycetes	Boliniales		
OTU198	Ascomycota	Dothideomycetes	Capnodiales		
OTU199	Basidiomycota	Agaricomycetes			
OTU205	Basidiomycota	Microbotryomycetes	Sporidiobolales	Sporidiobolales Incertae Sedis	Sporobolomyces
OTU207	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetaceae	Candida
OTU208	Ascomycota	Leotiomycetes	Helotiales	Sclerotiniaceae	Encoelia
OTU212	Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	Talaromyces
OTU215	Ascomycota	Leotiomycetes	Erysiphales	Erysiphaceae	Erysiphe
OTU217	Ascomycota	Dothideomycetes	Pleosporales		
OTU218	Basidiomycota	Agaricomycetes	Agaricales	Niaceae	Merismodes
OTU219	Basidiomycota	Exobasidiomycetes	Georgefischeriales	Tilletiariaceae	Tilletiaria
OTU223	Ascomycota	Lecanoromycetes	Teloschistales	Teloschistaceae	Xanthoria
OTU229	Ascomycota	Dothideomycetes	Pleosporales		
OTU231	Basidiomycota	Tremellomycetes	Tremellales	Tremellales Incertae Sedis	Holtermanniella
OTU233	Ascomycota	Sordariomycetes	Hypocreales	Ophiocordycipitaceae	Ophiocordyceps
OTU236	Ascomycota	Pezizomycetes	Pezizales	Ascobolaceae	
OTU243	Ascomycota	Sordariomycetes	Diaporthales	Valsaceae	Diaporthe
OTU245	Ascomycota	Pezizomycetes	Pezizales	Ascobolaceae	
OTU247	Basidiomycota	Exobasidiomycetes	Malasseziales	Malasseziaceae	Malassezia
OTU250	Ascomycota	Dothideomycetes	Pleosporales		
OTU251	Basidiomycota	Tremellomycetes	Tremellales	Tremellaceae	Cryptococcus
OTU252	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetaceae	Saccharomyces
OTU255	Ascomycota	Lecanoromycetes	Ostropales	Stictidaceae	Stictis

OTU257	Basidiomycota	Microbotryomycetes	Sporidiobolales	Sporidiobolales Incertae Sedis	Rhodotorula
OTU260	Ascomycota	Sordariomycetes	Sordariales	Chaetosphaeriaceae	Chaetosphaeria
OTU264	Basidiomycota	Agaricomycetes	Trechisporales	Hydnodontaceae	Trechispora
OTU268	Ascomycota	Leotiomycetes	Erysiphales	Erysiphaceae	Erysiphe
OTU270	Ascomycota	Leotiomycetes	Helotiales	Helotiaceae	Ascocoryne
OTU272	Ascomycota	Eurotiomycetes	Onygenales	Ascosphaeraceae	Ascosphaera
OTU276	Ascomycota	Sordariomycetes	Microascales	Microascales Incertae Sedis	Sphaeronaemella
OTU280	Ascomycota	Dothideomycetes	Pleosporales		
OTU284	Basidiomycota	Agaricostilbomycetes	Agaricostilbales	Agaricostilbaceae	Bensingtonia
OTU298	Ascomycota	Leotiomycetes	Helotiales	Helotiaceae	Hymenoscyphus
OTU299	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	Fusarium
OTU307	Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	Hormonema
OTU309	Ascomycota	Sordariomycetes			
OTU311	Ascomycota	Dothideomycetes	Capnodiales		
OTU312	Ascomycota	Sordariomycetes	Sordariomycetes Incertae Sedis		
OTU314	Chytridiomycota	Chytridiomycetes	Chytridiales		
OTU319	Ascomycota	Sordariomycetes	Sordariales		
OTU323	Ascomycota	Sordariomycetes	Hypocreales		
OTU328	Basidiomycota	Tremellomycetes	Tremellales		
OTU337	Basidiomycota	Agaricomycetes	Russulales	Bondarzewiaceae	Heterobasidion
OTU341	Basidiomycota	Agaricomycetes	Polyporales	Cystostereaceae	Cystostereum
OTU342	Ascomycota	Dothideomycetes	Pleosporales	Coniothyriaceae	Coniothyrium
OTU345	Basidiomycota	Agaricomycetes			
OTU349	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Acremonium
OTU350	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Geosmithia
OTU353	Ascomycota	Sordariomycetes	Sordariomycetes Incertae Sedis	Sordariomycetes Incertae Sedis	Xylochrysis
OTU357	Ascomycota	Sordariomycetes	Microascales	Microascaceae	Brachyconidiellopsis
OTU366	Ascomycota	Dothideomycetes	Pleosporales	Phaeosphaeriaceae	Saccharicola
OTU374	Ascomycota	Sordariomycetes	Glomerellales	Glomerellaceae	Colletotrichum
OTU379	Ascomycota	Dothideomycetes	Pleosporales	Phaeosphaeriaceae	Phaeosphaeria
OTU380	Basidiomycota	Agaricomycetes			
OTU385	Basidiomycota	Agaricomycetes	Polyporales	Ganodermataceae	Ganoderma
OTU386	Ascomycota	Sordariomycetes	Xylariales	Amphisphaeriaceae	Pestalotiopsis
OTU388	Basidiomycota	Agaricomycetes	Agaricales	Lyophyllaceae	Termitomyces
OTU391	Ascomycota	Leotiomycetes	Helotiales	Hemiphacidiaceae	Meria
OTU395	Ascomycota	Leotiomycetes	Thelebolales		
OTU422	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	
OTU423	Ascomycota	Sordariomycetes	Xylariales	Xylariaceae	Biscogniauxia
OTU426	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales		
OTU442	Ascomycota	Sordariomycetes	Hypocreales	Hypocreaceae	Trichoderma
OTU443	Ascomycota	Sordariomycetes	Hypocreales	Clavicipitaceae	Metarhizium
OTU446	Basidiomycota	Microbotryomycetes	Sporidiobolales	Sporidiobolales Incertae Sedis	Sporobolomyces
OTU447	Ascomycota	Leotiomycetes	Helotiales		

OTU456	Ascomycota	Sordariomycetes	Phyllachorales	Phyllachoraceae	Phyllachora
OTU461	Ascomycota	Dothideomycetes	Pleosporales		
OTU468	Zygomycota	Mucoromycotina	Mucorales	Pilobolaceae	Pilobolus
OTU471	Ascomycota	Sordariomycetes	Hypocreales		
OTU474	Ascomycota	Dothideomycetes	Botryosphaeriales	Botryosphaeriaceae	Neofusicoccum
OTU521	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	
OTU559	Ascomycota	Sordariomycetes	Coniochaetales	Coniochaetaceae	Lecythophora
OTU561	Ascomycota	Sordariomycetes	Sordariales	Sordariales Incertae Sedis	Remersonia
OTU562	Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	Hyaloscypha
OTU568	Basidiomycota	Tremellomycetes	Tremellales		
OTU581	Ascomycota	Dothideomycetes	Capnodiales		
OTU586	Basidiomycota	Cystobasidiomycetes	Erythrobasidiales	Erythrobasidiaceae	Erythrobasidium
OTU611	Ascomycota	Leotiomycetes	Helotiales		
OTU613	Ascomycota	Sordariomycetes	Sordariales	Sordariales Incertae Sedis	Remersonia
OTU642	Ascomycota	Saccharomycetes	Saccharomycetales	Debaryomycetaceae	Scheffersomyces
OTU670	Zygomycota	Mucoromycotina	Mucorales	Pilobolaceae	Pilobolus
OTU705	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Acremonium
OTU714	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	
OTU768	Ascomycota	Dothideomycetes	Capnodiales		
OTU808	Ascomycota	Sordariomycetes	Hypocreales		
OTU814	Ascomycota	Sordariomycetes	Sordariomycetes Incertae Sedis	Glomerellaceae	Colletotrichum
OTU837	Ascomycota	Sordariomycetes	Boliniales		
OTU838	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Acremonium
OTU856	Ascomycota	Sordariomycetes	Diaporthales	Gnomoniaceae	Asteroma
OTU858	Ascomycota	Sordariomycetes	Sordariomycetes Incertae Sedis	Plectosphaerellaceae	Lectera
OTU862	Ascomycota	Sordariomycetes	Sordariales		
OTU880	Ascomycota	Taphrinomycetes	Taphrinales	Protomycetaceae	Protomyces
OTU890	Ascomycota	Sordariomycetes	Hypocreales		
OTU920	Ascomycota	Leotiomycetes	Erysiphales	Erysiphaceae	Podosphaera
OTU931	Ascomycota	Leotiomycetes	Helotiales	Sclerotiniaceae	Sclerotinia
OTU965	Ascomycota	Dothideomycetes	Pleosporales		
OTU1014	Ascomycota	Sordariomycetes	Calosphaeriales	Pleurostomataceae	Pleurostomophora
OTU1065	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Acremonium
OTU1086	Ascomycota	Dothideomycetes	Pleosporales		
OTU1124	Ascomycota	Sordariomycetes	Hypocreales	Hypocreales Incertae Sedis	Acremonium
OTU1128	Ascomycota	Leotiomycetes	Thelebolales		
OTU1414	Ascomycota	Leotiomycetes	Thelebolales		
OTU1435	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	
OTU1445	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	
OTU1596	Basidiomycota	Tremellomycetes	Filobasidiales	Filobasidiaceae	
OTU1653	Ascomycota	Sordariomycetes	Sordariales	Sordariaceae	Neurospora
OTU1719	Ascomycota	Dothideomycetes	Pleosporales	Leptosphaeriaceae	Leptosphaeria

OTU1842	Basidiomycota	Pucciniomycetes	Pucciniales	Pucciniaceae	Puccinia
OTU1929	Basidiomycota	Agaricomycetes	Agaricales	Lyophyllaceae	Hypsizygus
OTU1954	Ascomycota	Eurotiomycetes	Chaetothyriales		
OTU2116	Ascomycota	Leotiomycetes	Helotiales		
OTU2151	Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	Cystofilobasidium
OTU2159	Ascomycota	Sordariomycetes	Sordariomycetes Incertae Sedis	Plectosphaerellaceae	Plectosphaerella
OTU2165	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetaceae	Kazachstania
OTU2288	Ascomycota	Dothideomycetes	Acrospermales	Acrospermaceae	Acrospermum
OTU2313	Basidiomycota	Pucciniomycetes	Pucciniales	Pucciniaceae	Puccinia

Appendix 3.6 Screenshot of phylogenetic tree zoomed in on Neocallimastigomycota branches.



Appendix 3.7 Negative correlations between FF OTUs and VOCs

Var1	Var2	Correlation
FOTU0	10.60_Ethyl Acetate	-0.72
FOTU40	10.60_Ethyl Acetate	-0.72
FOTU40	11.15_Methyl propionate	-0.78
FOTU44	11.15_Methyl propionate	-0.72
FOTU40	13.30_1-butanol	-0.86
FOTU0	13.30_1-butanol	-0.83
FOTU44	13.30_1-butanol	-0.82
FOTU7	13.30_1-butanol	-0.75
FOTU10	13.30_1-butanol	-0.74
FOTU2	13.30_1-butanol	-0.74
FOTU12	13.30_1-butanol	-0.74
FOTU0	14.30_n-Propyl acetate	-0.83
FOTU2	14.30_n-Propyl acetate	-0.79
FOTU12	14.30_n-Propyl acetate	-0.78
FOTU7	14.30_n-Propyl acetate	-0.78
FOTU44	14.30_n-Propyl acetate	-0.77
FOTU40	14.30_n-Propyl acetate	-0.76
FOTU40	14.63_Butanoic acid, methyl ester	-0.76
FOTU44	14.63_Butanoic acid, methyl ester	-0.72
FOTU40	16.50_Toluene	-0.83
FOTU44	16.50_Toluene	-0.74
FOTU40	16.92_2-Octene, (E)-	-0.75
FOTU0	17.61_Butanoic acid, ethyl ester	-0.74
FOTU40	17.61_Butanoic acid, ethyl ester	-0.72
FOTU2	17.61_Butanoic acid, ethyl ester	-0.72
FOTU12	17.61_Butanoic acid, ethyl ester	-0.71
FOTU0	17.99_Propanoic acid, propyl ester	-0.85
FOTU44	17.99_Propanoic acid, propyl ester	-0.81
FOTU40	17.99_Propanoic acid, propyl ester	-0.81
FOTU2	17.99_Propanoic acid, propyl ester	-0.79
FOTU7	17.99_Propanoic acid, propyl ester	-0.79
FOTU12	17.99_Propanoic acid, propyl ester	-0.79
FOTU10	17.99_Propanoic acid, propyl ester	-0.73
FOTU44	18.64_Pentanoic acid, methyl ester	-0.72
FOTU40	18.64_Pentanoic acid, methyl ester	-0.72
FOTU0	18.64_Pentanoic acid, methyl ester	-0.72
FOTU40	19.13_Butanoic acid, 1-methylethyl ester	-0.86
FOTU44	19.13_Butanoic acid, 1-methylethyl ester	-0.79
FOTU10	19.13_Butanoic acid, 1-methylethyl ester	-0.79
FOTU0	19.13_Butanoic acid, 1-methylethyl ester	-0.79
FOTU40	19.40_Butanoic acid	-0.74
FOTU40	19.45_Butanoic acid, 2-methyl-, ethyl ester	-0.80

FOTU44	19.45_Butanoic acid, 2-methyl-, ethyl ester	-0.74
FOTU0	19.45_Butanoic acid, 2-methyl-, ethyl ester	-0.74
FOTU0	19.65_Propanoic acid, 2-methyl-, propyl ester	-0.82
FOTU2	19.65_Propanoic acid, 2-methyl-, propyl ester	-0.79
FOTU40	19.65_Propanoic acid, 2-methyl-, propyl ester	-0.78
FOTU12	19.65_Propanoic acid, 2-methyl-, propyl ester	-0.78
FOTU44	19.65_Propanoic acid, 2-methyl-, propyl ester	-0.78
FOTU7	19.65_Propanoic acid, 2-methyl-, propyl ester	-0.77
FOTU40	19.67_Butanoic acid, 3-methyl-, ethyl ester	-0.78
FOTU0	19.67_Butanoic acid, 3-methyl-, ethyl ester	-0.75
FOTU44	19.67_Butanoic acid, 3-methyl-, ethyl ester	-0.74
FOTU2	19.67_Butanoic acid, 3-methyl-, ethyl ester	-0.71
FOTU12	19.67_Butanoic acid, 3-methyl-, ethyl ester	-0.70
FOTU9	20.60_Benzene, 1,3-dimethyl-	-0.75
FOTU0	20.94_1-Butanol, 2-methyl-, acetate	-0.71
FOTU40	21.44_Butanoic acid, 3-methyl-	-0.72
FOTU40	21.69_Propanoic acid, butyl ester	-0.92
FOTU0	21.69_Propanoic acid, butyl ester	-0.91
FOTU44	21.69_Propanoic acid, butyl ester	-0.89
FOTU10	21.69_Propanoic acid, butyl ester	-0.83
FOTU7	21.69_Propanoic acid, butyl ester	-0.81
FOTU12	21.69_Propanoic acid, butyl ester	-0.79
FOTU2	21.69_Propanoic acid, butyl ester	-0.79
FOTU40	22.05_1,7-Octadiene, 2,7-dimethyl-	-0.74
FOTU9	22.26_Heptanal	-0.73
FOTU40	22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.86
FOTU44	22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.74
FOTU10	22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.71
FOTU40	22.73_Pentanoic acid, 1-methylethyl ester	-0.76
FOTU0	22.73_Pentanoic acid, 1-methylethyl ester	-0.76
FOTU44	22.73_Pentanoic acid, 1-methylethyl ester	-0.74
FOTU44	23.22_Butanoic acid, 3-methyl-, propyl ester	-0.73
FOTU40	23.51_Methional	-0.76
FOTU0	23.51_Methional	-0.71
FOTU44	23.51_Methional	-0.71
FOTU0	23.58_Butanoic acid, 2-methylpropyl ester	-0.76
FOTU40	23.58_Butanoic acid, 2-methylpropyl ester	-0.72
FOTU40	25.23_1-Octen-3-ol	-0.70
FOTU9	26.01_Octanal	-0.74
FOTU40	27.14_Butanoic acid, 2-methylbutyl ester	-0.83
FOTU0	27.14_Butanoic acid, 2-methylbutyl ester	-0.81
FOTU44	27.14_Butanoic acid, 2-methylbutyl ester	-0.78
FOTU2	27.14_Butanoic acid, 2-methylbutyl ester	-0.74
FOTU7	27.14_Butanoic acid, 2-methylbutyl ester	-0.74
FOTU12	27.14_Butanoic acid, 2-methylbutyl ester	-0.74

FOTU9	27.66_Cyclohexanone, 2,2,6-trimethyl-	-0.79
FOTU40	28.27_Pentanoic acid, butyl ester	-0.95
FOTU44	28.27_Pentanoic acid, butyl ester	-0.90
FOTU0	28.27_Pentanoic acid, butyl ester	-0.90
FOTU10	28.27_Pentanoic acid, butyl ester	-0.84
FOTU7	28.27_Pentanoic acid, butyl ester	-0.80
FOTU12	28.27_Pentanoic acid, butyl ester	-0.78
FOTU2	28.27_Pentanoic acid, butyl ester	-0.77
FOTU9	29.51_Nonanal	-0.89
FOTU40	30.32_Pentanoic acid, 3-methylbutyl ester	-0.75
FOTU0	30.32_Pentanoic acid, 3-methylbutyl ester	-0.72
FOTU40	30.91_p-Cresol	-0.83
FOTU44	30.91_p-Cresol	-0.79
FOTU0	30.91_p-Cresol	-0.78
FOTU2	30.91_p-Cresol	-0.74
FOTU12	30.91_p-Cresol	-0.73
FOTU7	30.91_p-Cresol	-0.73
FOTU40	31.21_Phenylethyl Alcohol	-0.78
FOTU44	31.21_Phenylethyl Alcohol	-0.73
FOTU0	31.21_Phenylethyl Alcohol	-0.71
FOTU40	32.76_Decanal	-0.78
FOTU40	33.78_Phenol, 4-ethyl-	-0.81
FOTU44	33.78_Phenol, 4-ethyl-	-0.80
FOTU0	33.78_Phenol, 4-ethyl-	-0.79
FOTU2	33.78_Phenol, 4-ethyl-	-0.77
FOTU12	33.78_Phenol, 4-ethyl-	-0.75
FOTU7	33.78_Phenol, 4-ethyl-	-0.75
FOTU9	34.03_Tridecane	-0.75
FOTU9	35.46_2-Undecanone	-0.73
FOTU9	36.30_Dodecane, 2,6,10-trimethyl-	-0.76
FOTU40	38.57_Indole	-0.84
FOTU0	38.57_Indole	-0.81
FOTU44	38.57_Indole	-0.79
FOTU10	38.57_Indole	-0.74
FOTU7	38.57_Indole	-0.71

Appendix 3.8 Labels of OTUs in loadings for Pearson's correlation plot and circos plot (ITS1)

	Phylum	Class	Order	Family	Genus
OTU8	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	unknown_Herpotrichiellaceae
OTU19	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Ramularia
OTU40	Ascomycota	Sordariomycetes	Glomerellales	Plectosphaerellaceae	Plectosphaerella
OTU4	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU36	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	Neocallimastix
OTU35	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU30	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	Neocallimastigaceae
OTU31	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU10	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Zymoseptoria
OTU5	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU39	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU28	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU0	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Mycosphaerella
OTU11	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Ramularia
OTU12	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Exophiala
OTU15	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Ramularia
OTU45	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU9	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Ramularia
OTU22	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Ramularia
OTU37	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Exophiala
OTU29	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Veronaea
OTU1	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Septoria
OTU21	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	Piromyces
OTU13	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Sphaerulina
OTU27	Ascomycota	Sordariomycetes	unknown_Sordariomycetes	unknown_Sordariomycetes	unknown_Sordariomycetes
OTU7	Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Zymoseptoria

OTU16	Ascomycota	Dothideomycetes	Venturiales	Venturiaceae	Venturia
OTU18	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	Caecomyces
OTU26	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU23	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU38	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU20	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	Piromyces
OTU24	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknownNeocallimastigales	unknown_Neocallimastigales
OTU33	Ascomycota	Sordariomycetes	Glomerellales	Plectosphaerellaceae	Plectosphaerella
OTU25	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU34	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU17	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU32	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU42	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	Caecomyces
OTU2	Ascomycota	Dothideomycetes	Capnodiales	Cladosporiaceae	Cladosporium
OTU44	Ascomycota	Sordariomycetes	Glomerellales	Plectosphaerellaceae	Plectosphaerella
OTU14	Ascomycota	Leotiomycetes	Helotiales	unknownHelotiales	unknown_Helotiales
OTU43	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales
OTU41	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	unknown_Neocallimastigales	unknown_Neocallimastigales

Appendix 4.1 List of all compounds and their retention times identified in this study and scores for PC1 of PCAs in Figures 4.5-4.7.

Retention time and VOC	PC1
10.50_2-Butanone	0.09
18.25_Hexanal	0.08
25.44_Benzene, 1,2,4-trimethyl-	0.05
22.26_Heptanal	0.05
25.36_3-Octanone	0.05
14.21_Pentanal	0.05
16.66_Octane	0.05
25.14_1-Hepten-3-one	0.05
34.03_Tridecane	0.04
24.68_2-Heptanone, 5-methyl-	0.04
31.43_Undecane, 2,6-dimethyl-	0.04
35.85_Tridecane, 2-methyl-	0.04
20.60_Benzene, 1,3-dimethyl-	0.04
36.30_Dodecane, 2,6,10-trimethyl-	0.03
20.30_Ethylbenzene	0.03
24.37_2-Heptanone, 6-methyl-	0.03
32.95_Undecane, 3-methyl-	0.02
12.79_Butanal, 2-methyl-	0.02
27.70_Undecane	0.02
10.22_2,3-Butanedione	0.02
24.21_Decane	0.01
24.81_2-Heptenal, (Z)-	0.01
32.41_2-Decanone	0.01
26.63_Decane, 2-methyl-	0.00
29.81_Decane, 2,9-dimethyl-	0.00
13.88_2-Pentanone	0.00
35.46_2-Undecanone	0.00
8.98_Propanal, 2-methyl-	0.00
36.89_Tetradecane	0.00
25.55_5-Hepten-2-one, 6-methyl-	0.00
30.91_p-Cresol	-0.01
26.01_Octanal	-0.01
29.51_Nonanal	-0.01
29.99_Undecane, 2-methyl-	-0.01
24.74_Furan, 2-pentyl-	-0.02
13.43_Furan, 2-ethyl-	-0.02
13.80_1-Penten-3-one	-0.02
16.92_2-Octene, (E)-	-0.02
7.35_Acetone	-0.02
39.85_Pentadecane	-0.02
38.70_unknown alkane	-0.02
33.78_Phenol, 4-ethyl-	-0.03
25.66_2-Octanone	-0.03
7.19_Propanal	-0.03
6.62_Ethanol	-0.03
19.40_Butanoic acid	-0.03
24.54_: 2,6-Dimethyl-2-trans-6-octadiene	-0.03
26.20_D-Limonene	-0.04

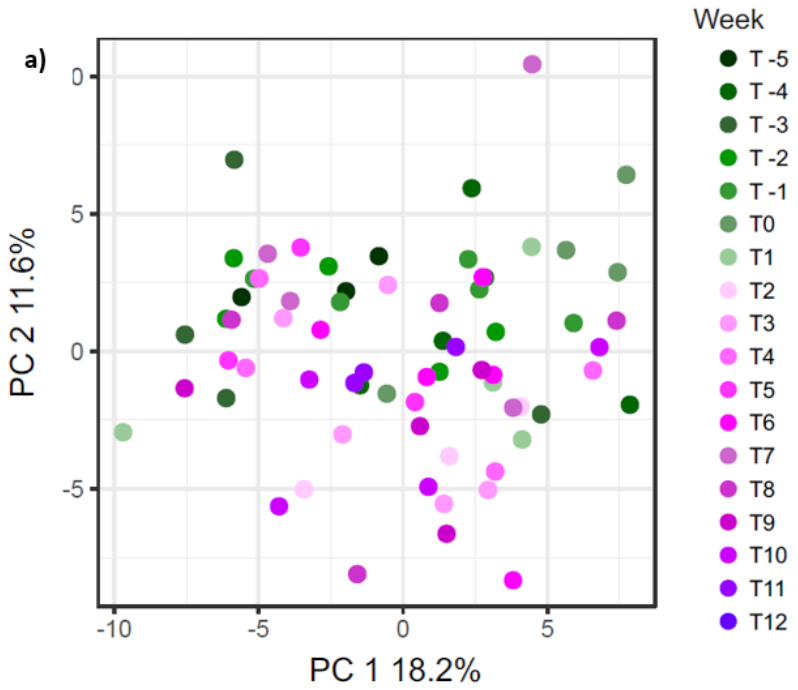
26.34_Hexanoic acid	-0.05
32.76_Decanal	-0.05
26.53_Dimethyl sulfone	-0.05
20.49_Nonane	-0.05
22.05_1,7-Octadiene, 2,7-dimethyl-	-0.05
37.80_2-Undecenal	-0.06
9.89_Furan, 2-methyl-	-0.06
22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	-0.07
10.80_2-Butanol	-0.07
23.27_2-Octene, 2,6-dimethyl-	-0.07
38.57_Indole	-0.07
29.16_2-Nonanone	-0.07
25.441_Benzaldehyde	-0.07
21.00_2-Hexenal	-0.08
17.27_1-Pentanol	-0.08
28.64_Benzeneacetaldehyde	-0.08
12.45_Butanal, 3-methyl-	-0.08
30.37_2(3H)-Furanone, 5-ethyldihydro-	-0.09
13.30_1-butanol	-0.09
29.00_Propanoic acid, hexyl ester	-0.09
11.15_Methyl propionate	-0.09
9.54_1-Propanol	-0.09
19.13_Butanoic acid, 1-methylethyl ester	-0.09
21.94_2-Heptanone	-0.10
16.50_Toluene	-0.11
10.60_Ethyl Acetate	-0.12
27.66_Cyclohexanone, 2,2,6-trimethyl-	-0.12
28.00_Phenol	-0.12
25.16_n-Caproic acid vinyl ester	-0.12
19.67_Butanoic acid, 3-methyl-, ethyl ester	-0.13
16.59_Isobutyl acetate	-0.13
19.45_Butanoic acid, 2-methyl-, ethyl ester	-0.13
22.97_Pentanoic acid	-0.13
24.12_cis-2,6-Dimethyl-2,6-octadiene	-0.14
19.65_Propanoic acid, 2-methyl-, propyl ester	-0.14
23.22_Butanoic acid, 3-methyl-, propyl ester	-0.14
15.91_Propanoic acid, 2-methyl-, ethyl ester	-0.15
14.63_Butanoic acid, methyl ester	-0.15
12.55_Acetic acid	-0.15
14.10_Propanoic acid, ethyl ester	-0.15
18.64_Methyl valerate	-0.16
18.36_Propanoic acid, 2-methyl-	-0.16
16.07_Propanoic acid	-0.16
20.83_1-Butanol, 3-methyl-, acetate	-0.17
20.94_1-Butanol, 2-methyl-, acetate	-0.17
14.30_n-Propyl acetate	-0.17
28.27_Pentanoic acid, butyl ester	-0.17
20.17_Propanoic acid, 2-methylpropyl ester	-0.17
17.61_Butanoic acid, ethyl ester	-0.18
21.44_Butanoic acid, 3-methyl-	-0.18
17.99_Propanoic acid, propyl ester	-0.18
21.73_Butanoic acid, 2-methyl-	-0.18
25.33_1-Butanol, 3-methyl-, propanoate	-0.19
21.69_Propanoic acid, butyl ester	-0.19
27.14_Butanoic acid, 2-methylbutyl ester	-0.19

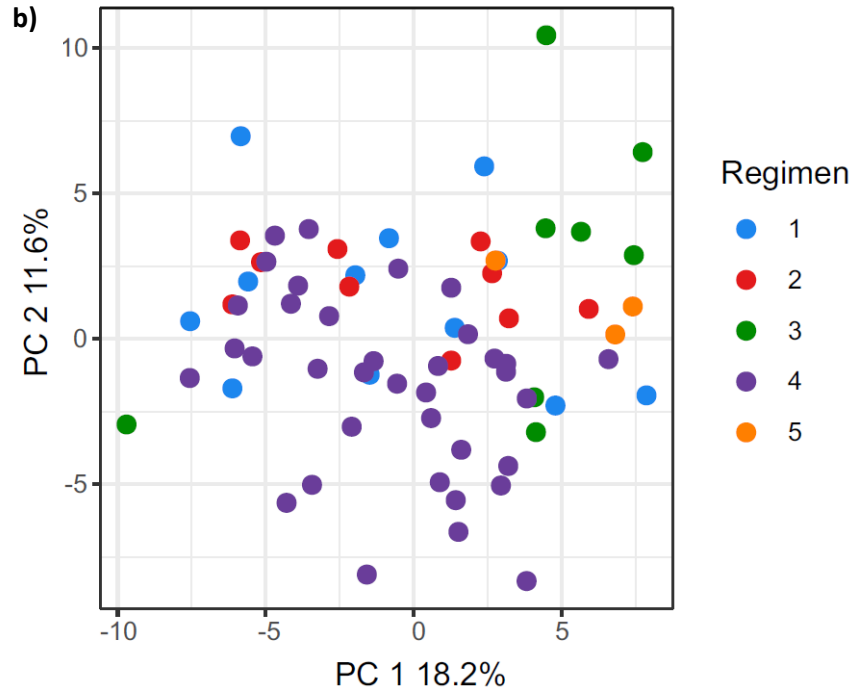
24.93_Pentanoic acid, propyl ester	-0.19
23.58_Butanoic acid, 2-methylpropyl ester	-0.20

Appendix 4.2 LME results of T0 and drug samples removed from analysis (factors time and regimen). P-values were adjusted using FDR.

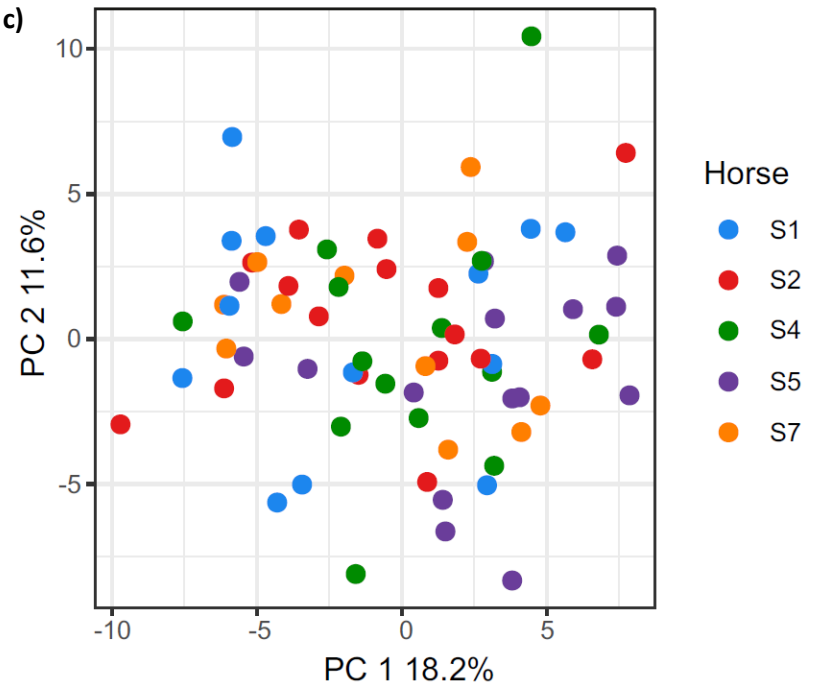
VOC	*p*-value	Adjusted *p*-value
Time		
Dimethyl sulfone	0.000	0.000
2-Octene, 2,6-dimethyl-	0.006	0.336
Nonane	0.012	0.420
Pentadecane	0.016	0.431
2,6-Dimethyl-2-trans-6-octadiene	0.029	0.584
Tridecane, 2-methyl-	0.039	0.584
cis-2,6-Dimethyl-2,6-octadiene	0.039	0.584
Regime		
Dimethyl sulfone	0.001	0.095
Propanoic acid, propyl ester	0.014	0.395
Furan, 2-pentyl-	0.016	0.395
1-Butanol, 3-methyl-, acetate	0.018	0.395
Hexanal	0.024	0.395
Propanoic acid, 2-methyl-, propyl ester	0.027	0.395
Propanoic acid, 2-methylpropyl ester	0.031	0.395
2-Octene, 2,6-dimethyl-	0.032	0.395
Butanoic acid, ethyl ester	0.036	0.395
Butanoic acid, 2-methylpropyl ester	0.038	0.395

Appendix 4.3 PCA results of T0 and AH samples removed from analysis. Points are coloured for the factors time in a) regimen in b) and horse in c).





c)

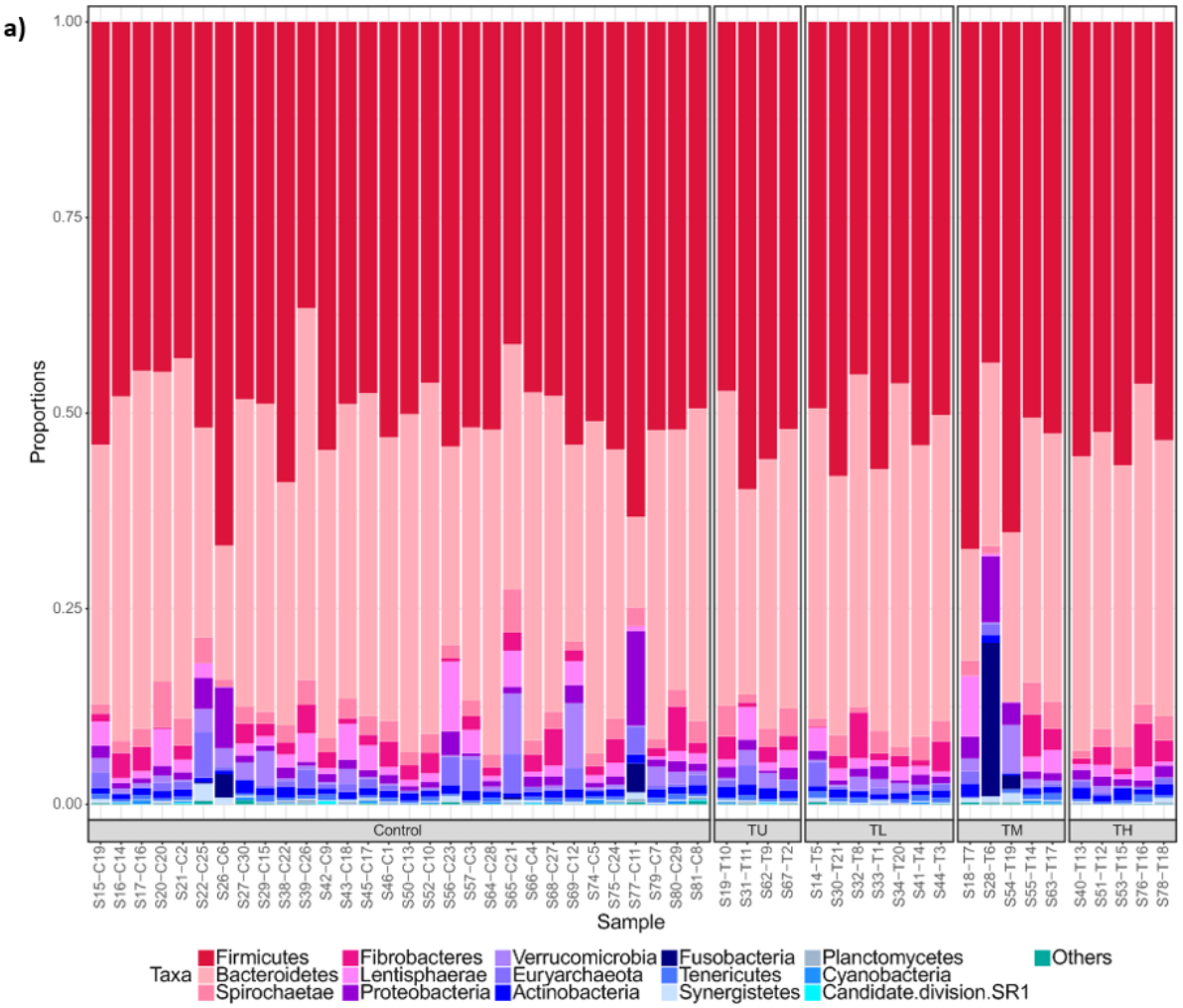


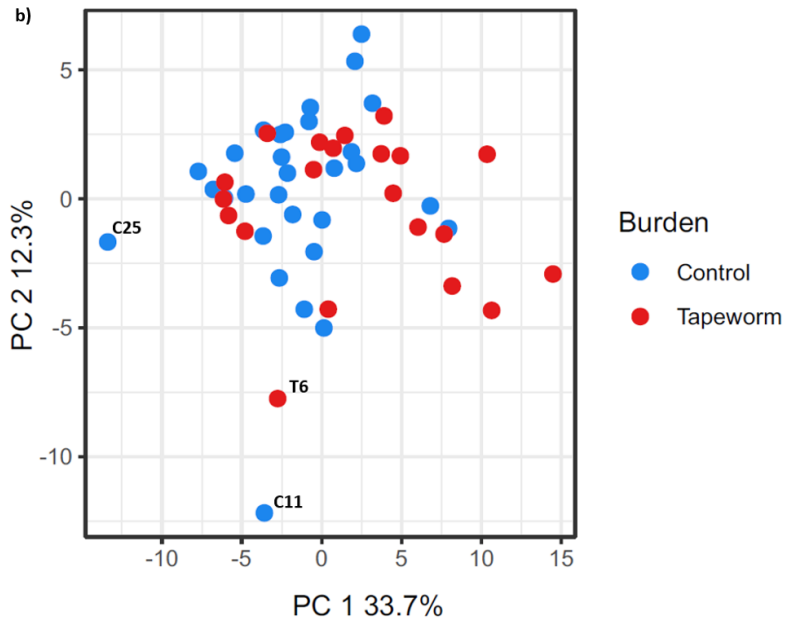
Appendix 5.1 The set of barcoded index primers.

Sample	Primer set	Forward	Reverse
Neg	N502_N701	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTC
T5	N502_N702	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTC
C19	N502_N703	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTC
C14	N502_N704	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTC
C16	N502_N705	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTC
T7	N502_N706	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T10	N502_N707	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
C20	N502_N708	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
C2	N502_N709	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
C25	N502_N710	AATGATACGGCGACCACCGAGATCTACACCTCTCTATACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
C6	N503_N702	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTC
C30	N503_N703	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTC
T6	N503_N704	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTC
C15	N503_N705	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTC
T21	N503_N706	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T11	N503_N707	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
T8	N503_N708	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
T1	N503_N709	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
T20	N503_N710	AATGATACGGCGACCACCGAGATCTACACTATCCTCTACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
C22	N504_N702	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTC
C26	N504_N703	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTC
T13	N504_N704	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTC
T4	N504_N705	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTC
C9	N504_N706	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
Sample	Primer set	Forward	Reverse
C18	N504_N707	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC

T3	N504_N708	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
C17	N504_N709	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
C1	N504_N710	AATGATACGGCGACCACCGAGATCTACACAGAGTAGAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTGACTGGAGTTCAGACGTGTGCTC
C13	N505_N702	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTC
T12	N505_N703	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTC
C10	N505_N704	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTC
T15	N505_N705	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTC
T19	N505_N706	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T14	N505_N707	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
C23	N505_N708	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
C3	N505_N709	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
T9	N506_N702	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTC
T17	N506_N703	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTC
C28	N506_N704	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTC
C21	N506_N705	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTC
C4	N506_N706	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
T2	N506_N707	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
C27	N506_N708	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
C12	N506_N709	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC
C5	N507_N702	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTC
C24	N507_N703	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTC
T16	N507_N704	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTC
C11	N507_N705	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTGACTGGAGTTCAGACGTGTGCTC
T18	N507_N706	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTGACTGGAGTTCAGACGTGTGCTC
C7	N507_N707	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTGACTGGAGTTCAGACGTGTGCTC
C29	N507_N708	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTGACTGGAGTTCAGACGTGTGCTC
C8	N507_N709	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTACACGACGCTC	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTGACTGGAGTTCAGACGTGTGCTC

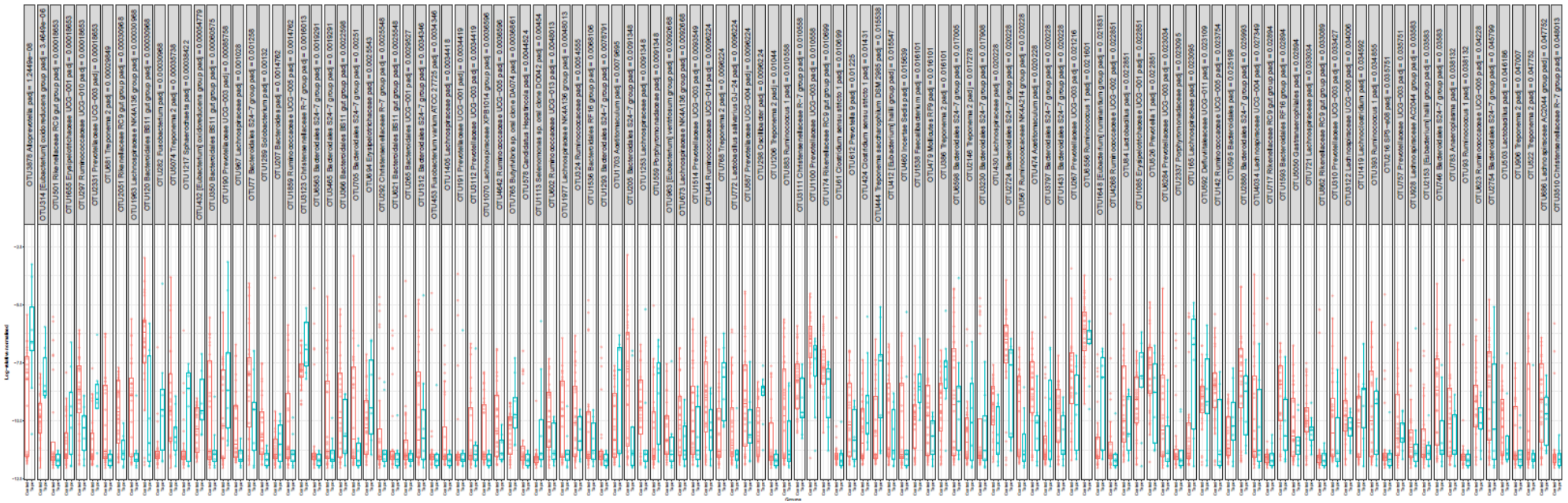
Appendix 5.2 In **a** is a taxonomic summary of all samples including outliers (T6, C11 and C25). In **b** a PCA of the VOC metabolome of all samples (with outliers) is shown.





Appendix 5.3 Significant OTUs comparing AT and CO.

Appendix 5.4 Significant OTUs comparing MH and CO.



Appendix 5.5 Significant OTUs comparing LSC and HSC.

Appendix 5.6 A list of the total VOCs identified and frequencies of occurrence in AT and CO samples.

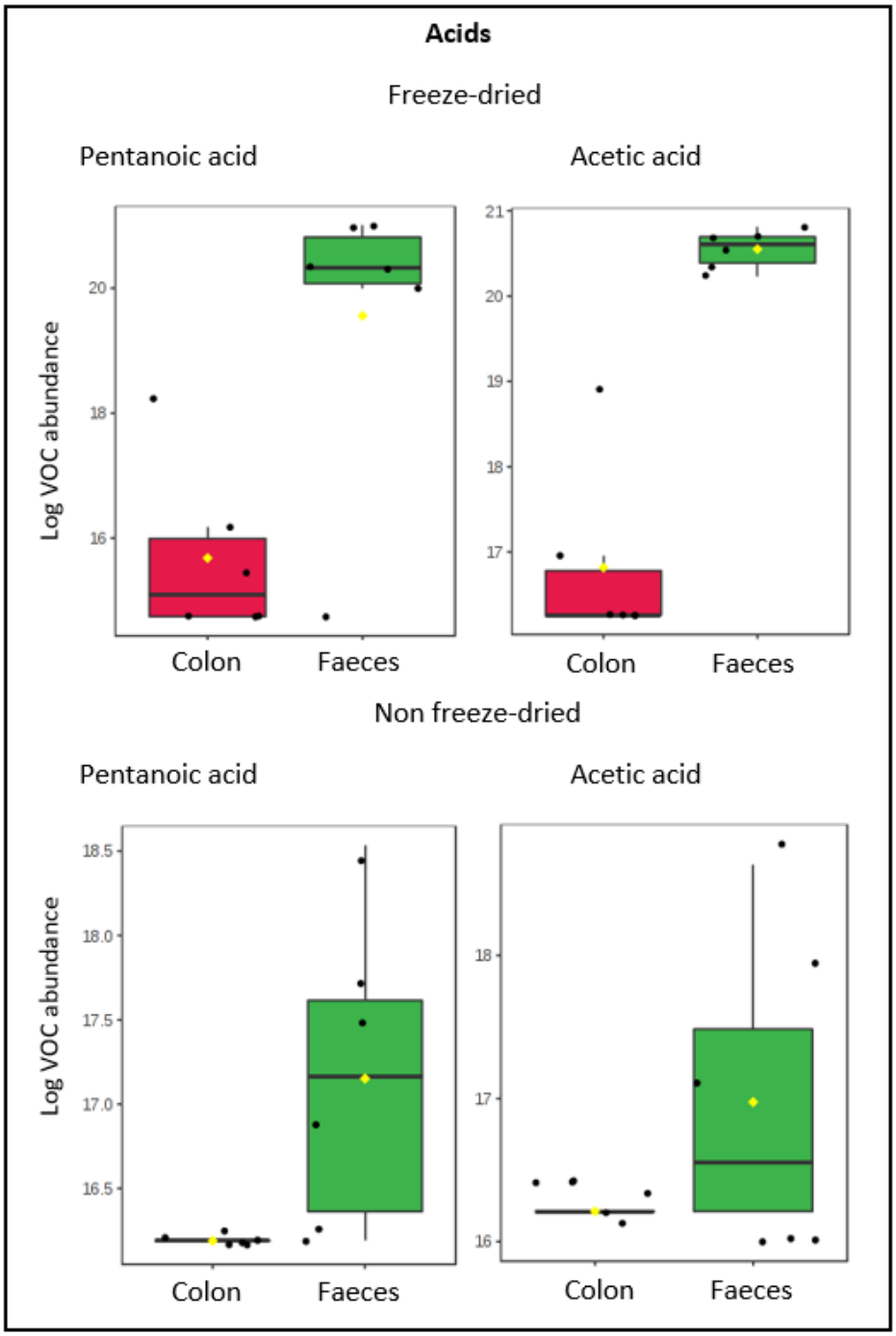
Retention time_VOC	Present Control	Present Tape
10.22_2,3-Butanedione	28	20
10.50_2-Butanone	21	20
10.60_Ethyl acetate	23	13
10.80_2-Butanol	24	17
11.93_1-Propanol, 2-methyl-	27	17
12.45_Butanal, 3-methyl-	28	20
12.55_Acetic acid	11	14
12.79_Butanal, 2-methyl-	28	20
13.43_Furan, 2-ethyl-	20	16
13.80_1-Penten-3-one	5	10
13.88_2-Pentanone	21	16
13.94_1-Penten-3-ol	28	18
14.21_Pentanal	28	20
14.30_n-Propyl acetate	16	15
14.63_Butanoic acid, methyl ester	25	17
15.21_Heptane, 2-methyl-	12	15
16.07_Propanoic acid	19	17
16.19_3-Pentanone, 2-methyl-	11	13
16.50_Toluene	25	19
16.66_Octane	27	19
16.92_2-Octene, (E)-	4	10
16.93_1-Propanone, 1-cyclopropyl-	26	17
17.27_1-Pentanol	26	19
17.61_Butanoic acid, ethyl ester	24	16
18.01_2-Hexanone	23	17
18.25_Hexanal	28	20
18.64_Methyl valerate	13	11
19.40_Butanoic acid	25	19
20.30_Ethylbenzene	27	20
20.49_Nonane	28	20
20.60_Benzene, 1,3-dimethyl-	28	20
21.00_2-Hexenal	23	17
21.32_1-Hexanol	14	11
21.71_Styrene	28	19
21.73_Butanoic acid, 2-methyl-	6	9
21.94_2-Heptanone	28	20
22.05_1,7-Octadiene, 2,7-dimethyl-	25	16
22.26_Heptanal	28	20
22.52_1,6-Octadiene, 3,7-dimethyl-, (S)-	27	20
22.85_Nonane, 2-methyl-	22	17

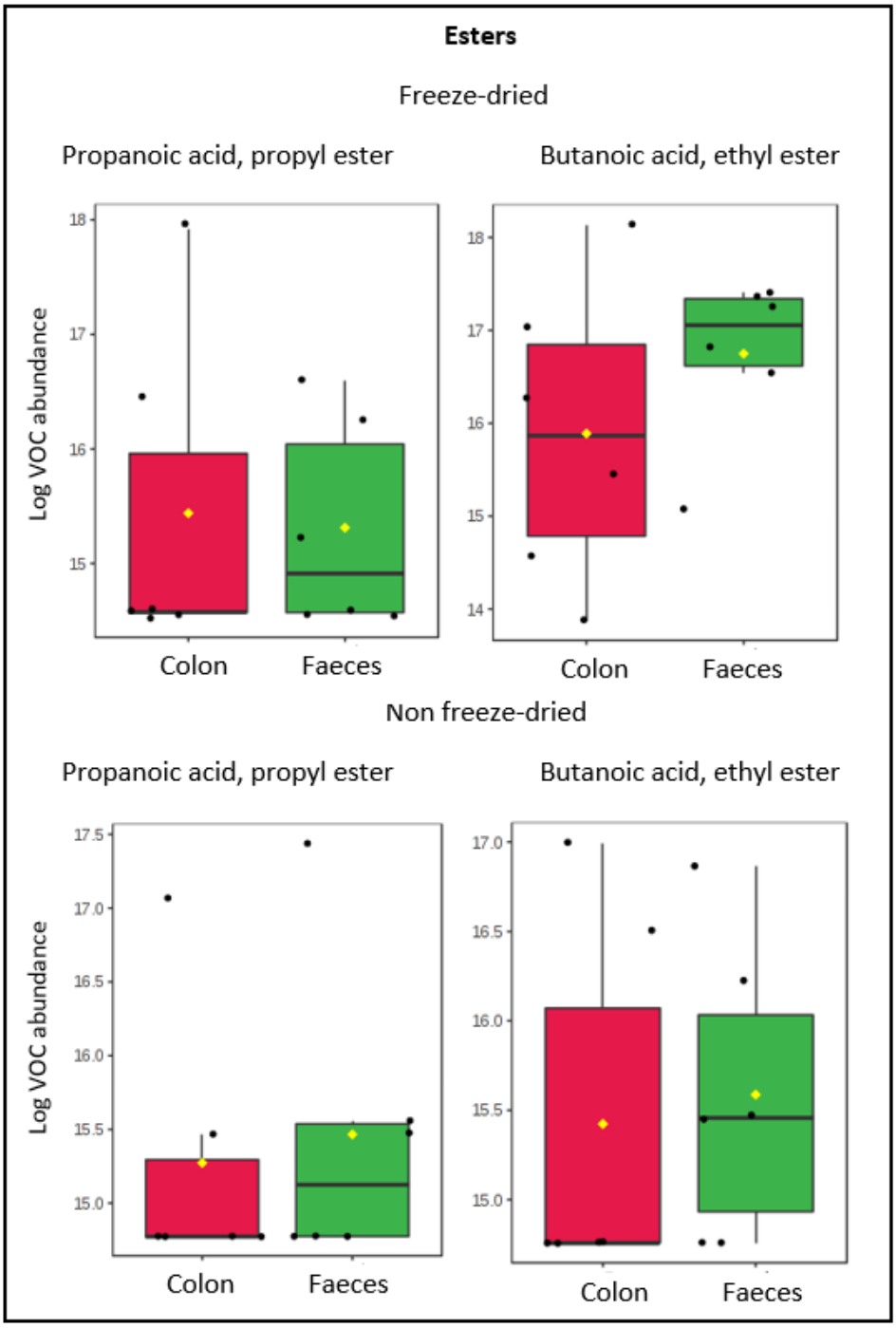
24.21_Decane	28	20
24.37_2-Heptanone, 6-methyl-	28	20
24.68_2-Heptanone, 5-methyl-	21	16
24.74_Furan, 2-pentyl-	28	19
24.90_1-Heptanol	5	11
25.23_1-Octen-3-ol	25	19
25.36_3-Octanone	27	20
25.44_Benzene, 1,2,4-trimethyl-	18	13
25.441_Benzaldehyde	15	10
25.55_5-Hepten-2-one, 6-methyl-	28	20
25.66_2-Octanone	12	14
25.68_3-Octanol	26	17
26.01_Octanal	28	20
26.20_D-Limonene	28	20
26.53_Dimethyl sulfone	28	20
27.66_Cyclohexanone, 2,2,6-trimethyl-	27	20
27.70_Undecane	15	9
28.00_Phenol	25	19
29.16_2-Nonanone	16	15
29.51_Nonanal	25	17
30.37_2(3H)-Furanone, 5-ethyldihydro-	20	14
30.75_Dodecane	27	20
30.91_p-Cresol	17	14
31.43_Undecane, 2,6-dimethyl-	27	20
31.50_Benzyl methyl ketone	9	9
32.41_2-Decanone	27	20
34.03_Tridecane	27	20
34.15_1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	28	20
35.46_2-Undecanone	22	18
35.73_1H-Pyrrole-2,5-dione, 3-ethyl-4-methyl-	9	11
36.01_Tridecane, 3-methyl-	21	16
36.30_Dodecane, 2,6,10-trimethyl-	28	20
36.89_Tetradecane	28	20
38.70_Heptadecane, 2,6,10,14-tetramethyl-	28	20
40.72_2-Butanone, 4-(2,6,6-trimethyl-1-cyclohexen-1-yl)-	8	10
6.62_Ethanol	28	20
7.19_Propanal	28	20
7.35_Acetone	28	20
7.48_Dimethyl sulfide	21	12
7.55_Isopropyl Alcohol	19	11
7.92_Acetic acid, methyl ester	28	20
8.98_Propanal, 2-methyl-	28	20
9.35_Methacrolein	6	10
9.54_1-Propanol	27	20

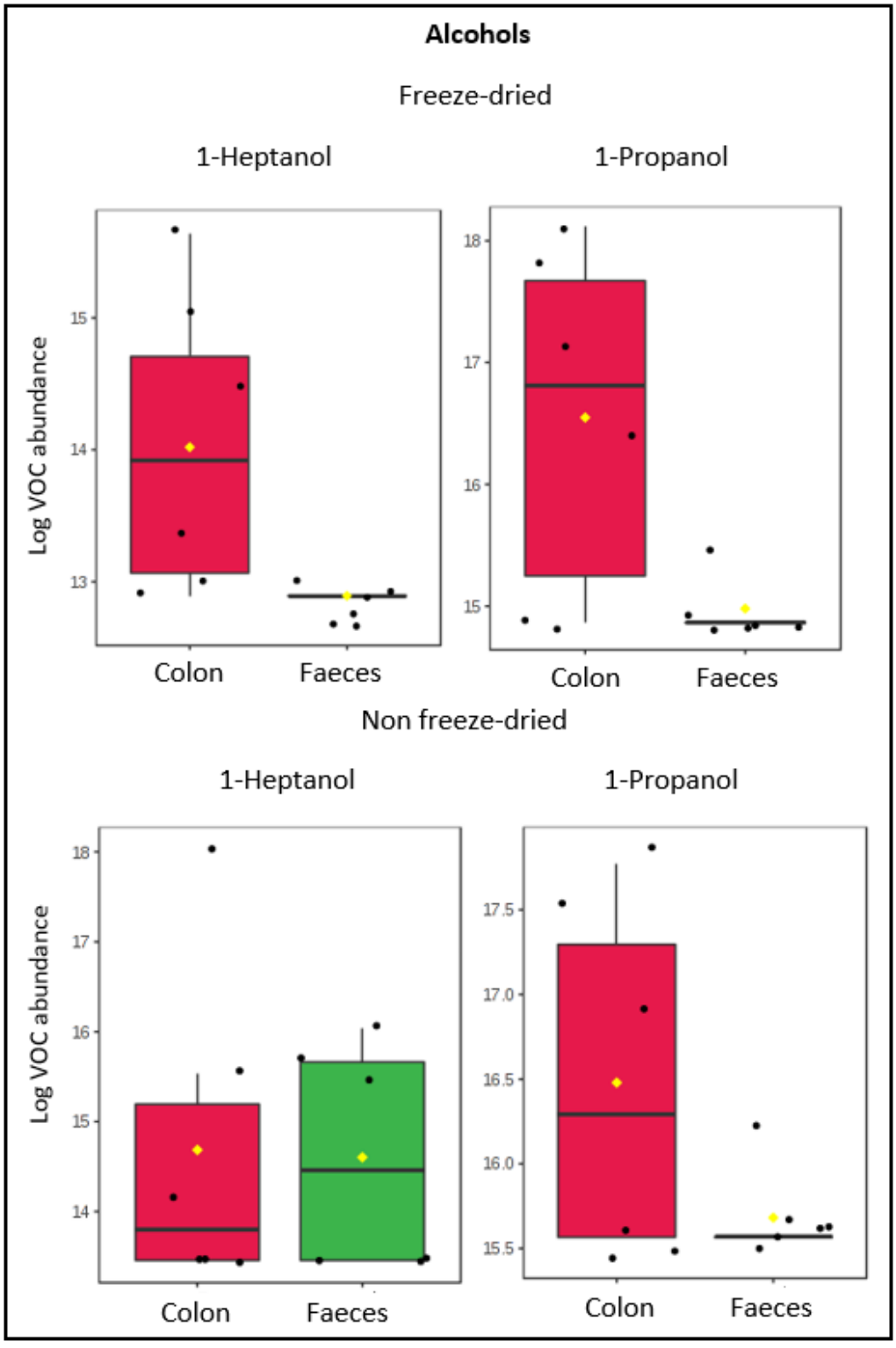
Appendix 5.7 Labels of OTUs in Figures 5.13a and 5.14c.

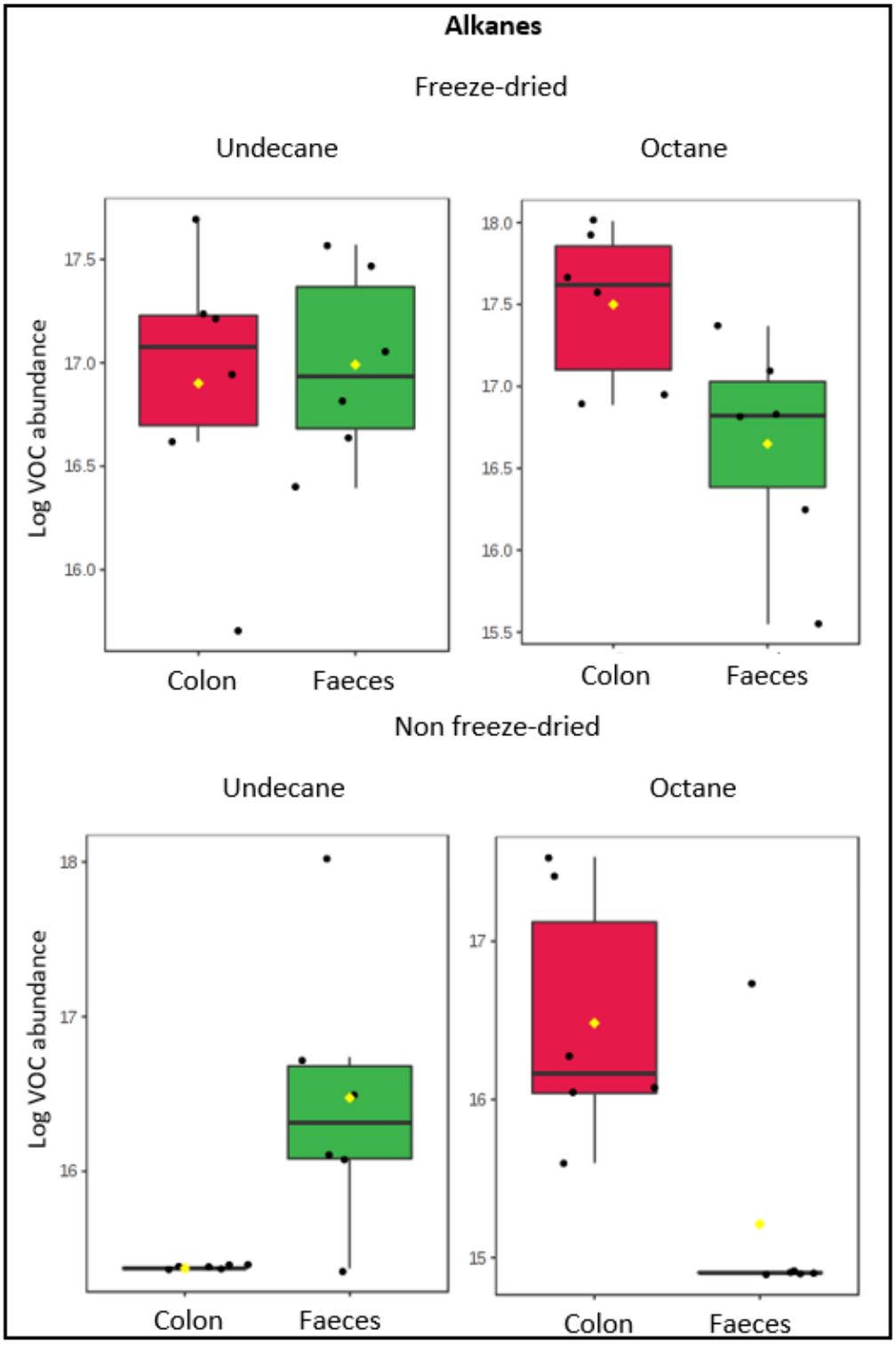
OTU	Kingdom	Phylum	Class	Order	Family
OTU13	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae
OTU14	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU102	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU111	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU142	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU147	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae
OTU185	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU235	Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae
OTU286	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU298	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU314	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU326	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU366	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidales BS11 gut group
OTU394	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU432	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU474	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU526	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae
OTU528	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU602	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU673	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU681	Bacteria	Spirochaetae	Spirochaetes	Spirochaetales	Spirochaetaceae
OTU691	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU705	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidales S24-7 group
OTU720	Bacteria	Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae
OTU724	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU743	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU768	Bacteria	Spirochaetae	Spirochaetes	Spirochaetales	Spirochaetaceae
OTU776	Bacteria	Firmicutes	Clostridia	Clostridiales	Christensenellaceae
OTU846	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU882	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU922	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU967	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU969	Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae
OTU974	Bacteria	Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae
OTU980	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU1127	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU1359	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU1396	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU1446	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU1508	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae
OTU1963	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU2051	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae
OTU2081	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU2243	Bacteria	Firmicutes	Clostridia	Clostridiales	Christensenellaceae
OTU2331	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae
OTU2508	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU2585	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidales S24-7 group
OTU3146	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU3205	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidales S24-7 group
OTU3249	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU3790	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
OTU3878	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae
OTU4038	Bacteria	Firmicutes	Negativicutes	Selenomonadales	Acidaminococcaceae
OTU4602	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae
OTU4857	Bacteria	Spirochaetae	Spirochaetes	Spirochaetales	Spirochaetaceae
OTU5074	Bacteria	Spirochaetae	Spirochaetes	Spirochaetales	Spirochaetaceae

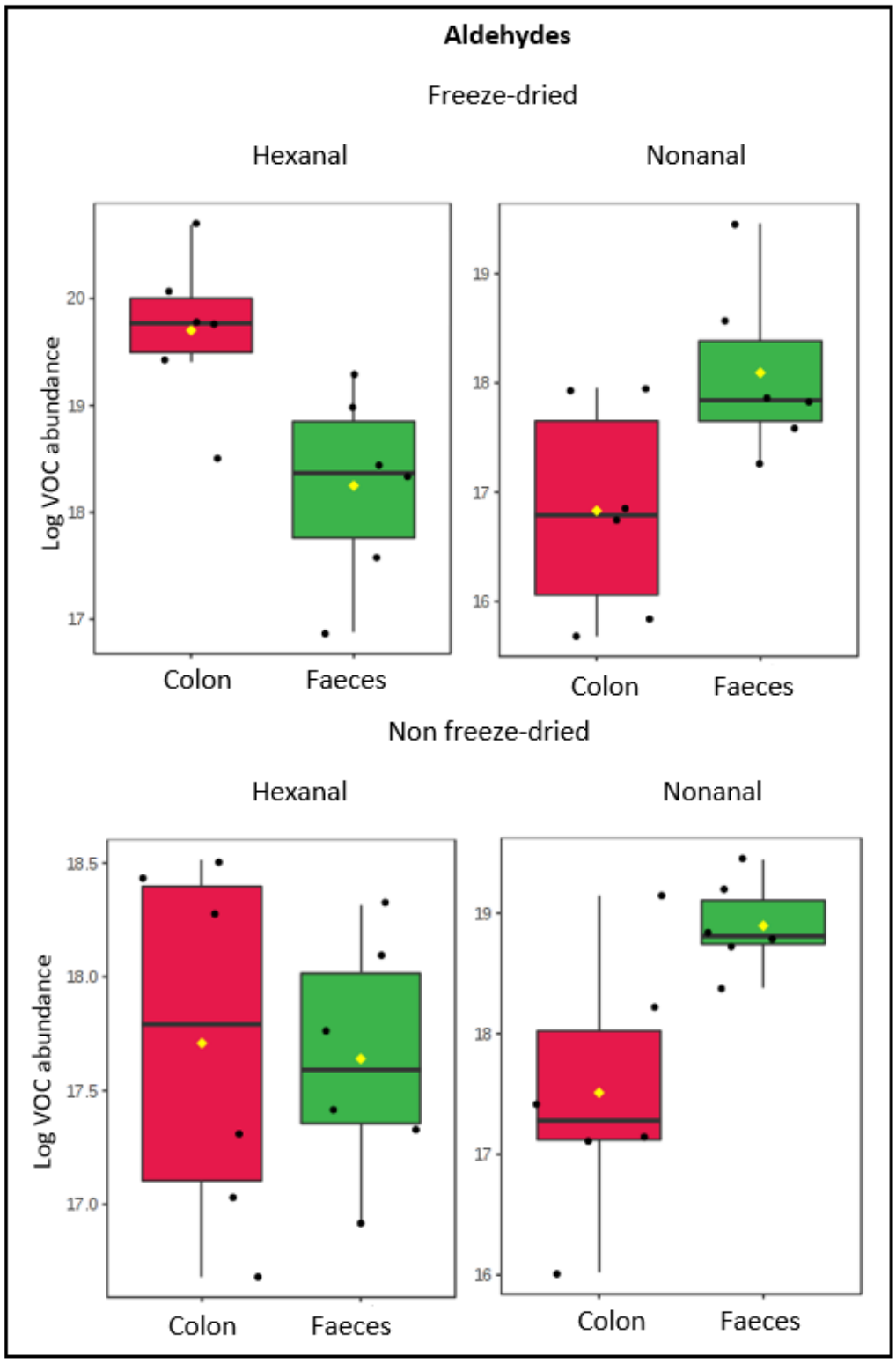
Appendix 6.1 Box plots of a selection of VOCs from key chemical classes to support Figure 6.4.




Esters
Freeze-dried
Propanoic acid, propyl ester
Butanoic acid, ethyl ester
Log VOC abundance
18
17
16
15
14
Colon
Faeces
Non freeze-dried
Propanoic acid, propyl ester
Butanoic acid, ethyl ester
Log VOC abundance
17.5
17.0
16.5
16.0
15.5
15.0
Colon
Faeces


Alcohols
Freeze-dried
1-Heptanol
1-Propanol
Log VOC abundance
Colon
Faeces
Non freeze-dried
1-Heptanol
1-Propanol
Log VOC abundance
Colon
Faeces


Alkanes
Freeze-dried
Undecane
Octane
Log VOC abundance
17.5
17.0
16.5
16.0
18.0
17.5
17.0
16.5
16.0
15.5
Colon
Faeces
Colon
Faeces
Non freeze-dried
Undecane
Octane
Log VOC abundance
18
17
16
17
16
15
Colon
Faeces
Colon
Faeces


Aldehydes
Freeze-dried
Hexanal
Nonanal
Log VOC abundance
20
19
18
17
Colon
Faeces
19
18
17
16
Colon
Faeces
Non freeze-dried
Hexanal
Nonanal
Log VOC abundance
18.5
18.0
17.5
17.0
Colon
Faeces
19
18
17
16
Colon
Faeces

Appendix 6.2a Table of VOC labels for component 1 (AT and CO).

RT	VOC	RT	VOC name
x38.91	Tetradecane, 3-methyl-	x22.97	Pentanoic acid
x25.36	3-Octanone	x12.45	Butanal, 3-methyl-
x28.64	Benzeneacetaldehyde	x34.15	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-
x14.21	3-Pentanone	x26.53	Dimethyl sulfone
x10.50	2-Butanone	x8.98	Propanal, 2-methyl-
x17.27	1-Pentanol	x31.43	Undecane, 2,6-dimethyl-
x18.25	Hexanal	x28.00	28.00_Phenol
x23.51	Methional	x36.30	Dodecane, 2,6,10-trimethyl-
x33.78	Phenol, 4-ethyl-	x38.57	Indole
x16.93	1-Propanone, 1-cyclopropyl-	x34.03	Tridecane
x20.30	Ethylbenzene	x29.81	Decane, 2,9-dimethyl-
x26.20	D-Limonene	x39.85	Pentadecane
x25.55	5-Hepten-2-one, 6-methyl-	x35.85	Tridecane, 2-methyl-
x20.60	Benzene, 1,3-dimethyl-	x17.61	Butanoic acid, ethyl ester
x29.16	2-Nonanone	x16.50	Toluene
x35.46	2-Undecanone	x12.55	Acetic acid
x21.44	Butanoic acid, 3-methyl-	x29.99	Undecane, 2-methyl-
x25.441	Benzaldehyde	x29.51	Nonanal
x32.76	Decanal	x19.40	Butanoic acid
x18.36	Propanoic acid, 2-methyl-	x21.71	Styrene
x24.12	cis-2,6-Dimethyl-2,6-octadiene	x32.41	2-Decanone
x36.01	Tridecane, 3-methyl-	x21.73	Butanoic acid, 2-methyl-
x10.22	2,3-Butanedione	x27.66	Cyclohexanone, 2,2,6-trimethyl-
x20.49	Nonane	x24.21	Decane
x24.37	2-Heptanone, 6-methyl-		
x22.52	1,6-Octadiene, 3,7-dimethyl-, (S)-		

Appendix 6.2b Table of VOC labels for components 2 and 3 (AT and CO).

Component 2		**Component 3**	
RT	VOC name	RT	VOC name
x33.78	Phenol, 4-ethyl-	x38.57	Indole
x17.27	1-Pentanol	x12.79	Butanal, 2-methyl-
x20.60	Benzene, 1,3-dimethyl-	x24.21	Decane
x20.30	Ethylbenzene	x29.16	2-Nonanone
x12.79	Butanal, 2-methyl-	x27.70	Undecane
x21.71	Styrene	x12.45	Butanal, 3-methyl-
x29.81	Decane, 2,9-dimethyl-	x38.70	Heptadecane, 2,6,10,14-tetramethyl-
x17.61	Butanoic acid, ethyl ester	x23.51	Methional
x24.12	cis-2,6-Dimethyl-2,6-octadiene	x13.80	1-Penten-3-one
x24.37	2-Heptanone, 6-methyl-	x21.00	2-Hexenal
x32.86	Dodecane, 2-methyl-	x38.91	Tetradecane, 3-methyl-
x28.00	Phenol	x20.49	Nonane
x35.85	Tridecane, 2-methyl-	x14.21	Pentanal
x10.50	2-Butanone	x26.01	Octanal
x23.51	Methional	x24.12	cis-2,6-Dimethyl-2,6-octadiene

Appendix 6.3a Table of VOC labels for component 1 and 2 (MH and CO).

Component 1		**Component 2**	
RT	VOC name	RT	VOC name
x17.27	1-Pentanol	x33.78	Phenol, 4-ethyl-
x32.76	Decanal	x16.50	Toluene
x38.91	Tetradecane, 3-methyl-	x12.79	Butanal, 2-methyl-
x20.30	Ethylbenzene	x20.49	Nonane
x22.05	1,7-Octadiene, 2,7-dimethyl-	x28.00	Phenol
x10.50	2-Butanone	x10.22	2,3-Butanedione
x24.37	2-Heptanone, 6-methyl-	x27.70	Undecane
x28.64	Benzeneacetaldehyde	x22.05	1,7-Octadiene, 2,7-dimethyl-
x34.15	1-Cyclohexene-1-carboxaldehyde, 2,6,6-trimethyl-	x14.21	Pentanal
x20.60	Benzene, 1,3-dimethyl-	x29.16	2-Nonanone
x16.92	2-Octene, (E)-		
x24.12	cis-2,6-Dimethyl-2,6-octadiene		
x25.36	3-Octanone		
x18.36	Propanoic acid, 2-methyl-		
x26.20	D-Limonene		
x27.66	Cyclohexanone, 2,2,6-trimethyl-		
x23.51	Methional		
x21.73	Butanoic acid, 2-methyl-		
x16.66	Octane		
x28.00	Phenol		
x21.44	Butanoic acid, 3-methyl-		
x26.01	Octanal		
x16.93	1-Propanone, 1-cyclopropyl-		
x16.50	Toluene		
x33.78	Phenol, 4-ethyl-		

Appendix 6.3b Table of VOC labels for component 3 (MH and CO).

Component 3			
RT	VOC name	RT	VOC name
x16.66	Octane	x12.45	Butanal, 3-methyl-
x14.21	Pentanal	x21.73	Styrene
x25.36	3-Octanone	x31.43	Undecane, 2,6-dimethyl-
x21.71	Styrene	x16.50	Toluene
x22.26	Heptanal	x21.00	2-Hexenal
x20.60	Benzene, 1,3-dimethyl-	x23.51	Methional
x17.61	Butanoic acid, ethyl ester	x29.16	2-Nonanone
x29.51	Nonanal	x16.92	2-Octene, (E)-
x37.19	2-Hexadecene, 3,7,11,15-tetramethyl-, [R-[R*,R*-(E)]]-		
x12.79	Butanal, 2-methyl-		
x10.60	Ethyl Acetate		
x20.30	Ethylbenzene		
x26.01	Octanal		
x35.46	2-Undecanone		
x13.88	2-Pentanone		
x32.86	Dodecane, 2-methyl-		
x17.27	1-Pentanol		
x13.80	2-Butanol		
x21.94	2-Heptanone		
x18.25	Hexanal		
x35.85	Tridecane, 2-methyl-		
x25.441	Benzaldehyde		
x19.65	Propanoic acid, 2-methyl-, propyl ester		
x24.74	Furan, 2-pentyl-		
x21.44	Butanoic acid, 3-methyl-		
x34.03	Tridecane		
x36.01	Tridecane, 3-methyl-		

General Appendix

Publication arising from **Chapter 2**

Metabolomics (2018) 14:19
https://doi.org/10.1007/s11306-017-1315-7

ORIGINAL ARTICLE

A comparison of sample preparation methods for extracting volatile organic compounds (VOCs) from equine faeces using HS-SPME

Rachael Hough[1] · Debra Archer[2] · Christopher Probert[1]

Received: 28 September 2017 / Accepted: 22 December 2017

Abstract
Introduction Disturbance to the hindgut microbiota can be detrimental to equine health. Metabolomics provides a robust approach to studying the functional aspect of hindgut microorganisms. Sample preparation is an important step towards achieving optimal results in the later stages of analysis. The preparation of samples is unique depending on the technique employed and the sample matrix to be analysed. Gas chromatography mass spectrometry (GCMS) is one of the most widely used platforms for the study of metabolomics and until now an optimised method has not been developed for equine faeces.
Objectives To compare a sample preparation method for extracting volatile organic compounds (VOCs) from equine faeces.
Methods Volatile organic compounds were determined by headspace solid phase microextraction gas chromatography mass spectrometry (HS-SPME-GCMS). Factors investigated were the mass of equine faeces, type of SPME fibre coating, vial volume and storage conditions.
Results The resultant method was unique to those developed for other species. Aliquots of 1000 or 2000 mg in 10 ml or 20 ml SPME headspace were optimal. From those tested, the extraction of VOCs should ideally be performed using a divinylbenzene-carboxen-polydimethysiloxane (DVB-CAR-PDMS) SPME fibre. Storage of faeces for up to 12 months at − 80 °C shared a greater percentage of VOCs with a fresh sample than the equivalent stored at − 20 °C.
Conclusions An optimised method for extracting VOCs from equine faeces using HS-SPME-GCMS has been developed and will act as a standard to enable comparisons between studies. This work has also highlighted storage conditions as an important factor to consider in experimental design for faecal metabolomics studies.

Keywords Volatile organic compounds · GC-MS · SPME · Faeces · Equine

1 Introduction

The equine hindgut contains a complex and diverse community of microorganisms (microbiota). The microbiota is essential for the breakdown of fibre into volatile fatty acids (VFA), providing a major source of energy (Bergman 1990). Disturbance to the hindgut microbiota can be detrimental to equine gastrointestinal health resulting in colic (Daly et al. 2012) and laminitis (Milinovich et al. 2007). Metabolomics provides a tool for characterising the functionality of the gut microbiota and therefore is a useful means for the study of gut dysbiosis (Marcobal et al. 2013).

Volatile organic compounds (VOCs) are a large group of carbon-containing molecules, which may be of biological or synthetic origin. The low molecular weight and high vapour pressure of VOCs allows them to enter the gaseous phase at room temperature; contributing to the odour of faeces, urine, breath, saliva, blood and sweat. These compounds may be generated by physiological processes from the host or by its microbiota (Amann et al. 2014). VOC analysis of faeces provides a simple approach to understanding functional changes of the microbiota of the distal intestine. The faecal VOC profiles of horses have been found to alter in response to dietary supplementation, from the ingestion of probiotics and when

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11306-017-1315-7) contains supplementary material, which is available to authorized users.

✉ Rachael Hough
rachael.hough@liverpool.ac.uk

[1] Department of Cellular and Molecular Physiology, University of Liverpool, Liverpool, UK

[2] Department of Epidemiology and Population Health, University of Liverpool, Liverpool, UK

Published online: 04 January 2018

suffering from gastrointestinal disease (colic) (Turner et al. 2013; Ishizaka et al. 2014; Proudman et al. 2014).

Headspace, solid phase microextraction gas chromatography mass spectrometry is a rapid and economical technique that has been used to characterise the faecal metabolome in human subjects and many species including ruminants, poultry and the horse (Garner et al. 2007, 2008; Stavert et al. 2014). The use of SPME to extract analytes from biological samples is advantageous as it does not require the addition of organic solvents (Arthur and Pawliszyn 1990). GCMS is one of the most widely used platforms for the study of metabolomics yet an optimised method has yet to be studied for equine faeces. The preparation steps of samples for metabolomics are vital for achieving optimal results and consistent data in the later stages and a standardised method will allow the comparison of results between laboratories. It has previously been demonstrated that vial volume, SPME fibre coating and mass of faecal material have an effect on the number and abundance of VOCs in human and murine faecal samples (Reade et al. 2014). This is not surprising as faecal VOC profiles differ between species (Saric et al. 2008). In addition, the optimal preparation steps differed between human and murine faeces (Reade et al. 2014). Therefore, it is expected that a modified method for GCMS VOC profiling of equine faeces is needed.

The sample preparation factors investigated in this work were: faecal sample mass, SPME fibre type, headspace vial volume and the reproducibility of the optimised method proposed. Sample storage for varying time and temperatures were also investigated in one individual pony.

2 Methods

2.1 Animals and sample collection

2.1.1 Part (A)

Faecal samples were collected in June 2015 from four horses immediately after spontaneous defaecation. The horses were Thoroughbred and Irish sport horse crossbreeds, demographic information for each horse is listed in Supplementary Table 1. Samples were transported on ice and transported to the laboratory where they were frozen at − 20 °C within 0.5–3 h and were stored until analysis. All horses were housed on the same premises (Phillip Leverhulme Equine Hospital, Leahurst Campus, University of Liverpool) and were maintained under the same conditions: living out at pasture all year round with a diet supplemented with hay during the winter. The horses were stabled for periods during the daytime to be handled by students. During this time horses had access to ad libitum hay. Each horse was treated for intestinal parasites with ivermectin and praziquantel 4 weeks previously.

2.1.2 Part (B)

A faecal sample was collected from a mixed-breed pony mare (P1) on private premises in May 2016 to investigate the effect of storage on faecal VOCs. The sample was collected after spontaneous defaecation and immediately transported to the laboratory (< 2 h). Samples were stored at either − 20 °C (n = 9) or − 80 °C (n = 9) until analysis. The pony (aged 16 years) was maintained at pasture all year round and received a diet supplemented with hay during the winter. At the time of collection, a faecal egg count of 0.0. eggs per gram was recorded and praziquantel had been administered 4 weeks previously. A summary of the collection and division of samples from Part A and Part B is shown in Table 1.

2.2 Head space-solid phase micro extraction (HS-SPME)

VOCs were extracted from the headspace of faeces when the SPME fibre was exposed to the headspace of a 10 ml or 20 ml glass vial (Supelco, Dorset, UK) as programmed using Combi Pal auto-sampler (CTC Analytics, Switzerland). Vials were incubated at 60 °C for 30 min, followed by a 20 min extraction at 60 °C. The SPME fibres used were divinylbenzene-carboxen-polydimethysiloxane (DVB-CAR-PDMS) 50/30 µm (1 cm) or carboxen-polydimethysiloxane (CAR-PDMS) (85 µm) (Sigma-Aldridge, Dorset, UK). The SPME fibres were pre-conditioned in accordance with manufacturer's instructions before use.

2.3 GC-MS conditions

GC-MS analysis was carried out using a Perkin Elmer Clarus 500 GC/MS Quadruple bench top system (Beaconsfield, UK). Separation of VOCs was performed in the Zebron ZB-624 GC column with an inner diameter of 0.25 mm, length of 60 m and a film thickness of 1.4 µm (Phenomenex, Macclesfield, UK). Helium of 99.996% purity was used as a carrier gas, set at a flow rate of 1 ml/min (BOC, Sheffield, UK). SPME fibre desorption temperature and time were 220 °C and 5 min, respectively. The GC oven was initially set at 40 °C and held for 1 min before a ramp to 220 °C at a rate of 5 °C per min and held at 4 min (total run time of 41 min). The MS was operated in electron impact ionization EL + mode, scanning ion mass fragments from 10 to 300 *m/z* with an inter-scan delay of 0.1 s and a resolution of 1000 FWHM (Full Width at Half Maximum). Laboratory air was sampled frequently to rule out possible contaminants from the analysis. Blank vials were tested between samples to ensure VOCs were originating from faeces and to prevent

Table 1 A faecal sample was collected from Horse 1, 2, 3, 4 and Pony 1 to compare methods for extracting VOCs using HS-SPME-GCMS

Experimental factor	Horse 1	Horse 2	Horse 3	Horse 4
	Number of technical replicates			
Sample mass				
100 mg	3	3	3	3
1000 mg	3*	3*	3*	3*
2000 mg	3	3	3	3
SPME fibre type				
DVB-CAR-PDMS	3*	3*	3*	3*
CAR-PDMS	3	3	3	3
Headspace vial volume				
10 ml	3*	3*	3*	3*
20 ml	3	3	3	3
	Pony 1			
Storage				
Fresh	3			
− 20 °C 1 week	3			
− 20 °C 6 months	3			
− 80 °C 1 week	3			
− 80 °C 6 months	3			

Experimental conditions and the numbers of technical replicates are listed. Technical replicates marked with * indicate these were included in an additional analysis to test the reproducibility of the method

a carry-over of VOCs on the SPME fibre between samples. Three commercially available standards of compounds putatively identified in this study (benzaldehyde, 2-pentanone and indole) were run and the data processed in the same way as experimental data to validate compound identification towards MSI level 1 (Sumner et al. 2007). Internal standards were not used for this non-targeted study because of the limited knowledge of compounds emitted from the equine faecal metabolome.

2.4 Faecal mass optimisation

Triplicates of 100 mg (mean and SEM) (105.9 ± 1.3 mg), 1000 mg (1003.0 ± 4.1 mg), 2000 mg (2005.8 ± 5.1 mg) of faeces from each horse from Part A were divided into 10 ml vials to determine the number and abundance of VOCs. The SPME fibre DVB-CAR-PDMS was used to extract VOCs before injection into the GC oven.

2.5 SPME fibre type

Two different fibre coatings were tested in order to determine whether the type of fibre coating used in SPME-GC-MS analysis has an effect on the VOC profile. The SPME fibre coatings chosen were CAR-PDMS and DVB-CAR-PDMS. Masses of 1000 mg (1008.9 ± 3.3 mg) were placed into six 10 ml vials for each horse from Part A. Three replicates were assigned to the CAR-PDMS fibre coating group and 3 replicates for the DVB-CAR-PDMS group. The samples then underwent HS-SPME-GC-MS analysis.

2.6 Headspace volume

Vials of 10 and 20 ml were chosen in order to compare 1000 mg of faeces. Three 10 ml and three 20 ml vials containing 1000 mg (mean and SEM, 1003.9 ± 2.3 mg) of faeces from each horse from Part A underwent HS-SPME-GC-MS analysis using a DVB-CAR-PDMS fibre to extract VOCs.

2.7 Technical replicates

A total of nine technical replicates of 1000 mg of faeces were analysed from horses (H1, H2, H3 and H4) from Part A. All samples were contained in 10 ml vials and VOCs were extracted using the SPME fibre DVB/CAR/PDMS. In order to assess the variation across technical replicates, statistical analysis was performed.

2.8 Time and temperature of storage

Aliquots of faeces (1008 ± 0.6 mg, mean and SEM) collected from pony P1 (Part B) were placed into headspace vials. The samples were stored for 1 week, 6 and 12 months at − 20 or − 80 °C before analysis. Three technical replicates of each storage condition were analysed. A DVB-CAR-PDMS fibre was used to extract VOCs before injection into the GC oven.

Three fresh aliquots of the faecal sample collected underwent HS-SPME-GCMS within 2 h of defecation to compare against frozen samples.

2.9 Data processing

The data was processed using Automated Mass Spectral Deconvolution System (AMDIS-version 2.71, 2012) and the National Institute of Standards and Technology (NIST) mass spectral library (version 2.0, 2011) to putatively identify VOCs. The R package Metab (Aggio et al. 2011) was used to align the data. All samples were analysed in triplicate, an average was taken of the replicates at this stage and taken forward for data analysis to keep random errors to a minimum. Statistical analysis was performed in R version 3.1.2 and using the online software tool Metaboanalyst 3.0 (Xia et al. 2012).The data was filtered by removing VOCs that were not present in 50% of samples within at least one of the experimental conditions being compared. Missing values were imputed with half minimum values from each data matrix. The abundance data were normalised using log transformation (general logarithm) and group means compared using an independent t-test or one-way ANOVA without interactions where appropriate. Pair-wise comparisons were made by Tukey's HSD test, followed by Bonferroni correction. Fisher's exact test, followed by Bonferroni correction was performed to assess for statistical significance in absence or presence of compounds between conditions. All p values of less than 0.05 were considered to be significant. Principal component analysis (PCA) and dendrograms (hierarchical clustering) were constructed to visually compare VOC profiles.

3 Results

3.1 Faecal mass

The mean (±SD) number of VOCs detected from each sample mass were 59 (±9.5), 79 (±8.8) and 80 (±4.8) for 100, 1000 and 2000 mg, respectively. There were 13 VOCs (Supplementary Table 2) that were exclusive to 1000 and 2000 mg. All compounds detected were present in at least one sample of 1000 mg whereas three compounds (butanoic acid, 3-methyl-, propyl ester, propanoic acid, 2-methyl-, acetic acid) were missing from 2000 mg samples. One compound (2-decanone), was significantly lower in abundance in 100 mg than 1000 and 2000 mg ($p < 0.01$, ANOVA, Tukey's HSD test and Bonferroni corrected). The coefficient of variation (CV) of VOC peak area was calculated for VOCs shared between replicated of each sample mass. Coefficient of variation values were 1.6–13, 1.1–14.4, 0.8–12.5% for 100, 1000 and 2000 mg, respectively. A full list if CV values for each VOC are shown in Supplementary Table 3. A PCA is shown in Fig. 1 which represents the VOC profiles of each sample mass. A list of scores for PC1 and PC2 is supplied in Supplementary Table 4.

3.2 SPME fibre type

Mean (±SD) number of compounds detected by the CAR-PDMS SPME fibre was 52 (±11.2) and DVB-CAR-PDMS was 78 (±7.8). Significantly more VOCs were detected with a DVB-CAR-PDMS fibre than from a CAR-PDMS fibre ($p < 0.01$, t test, Bonferroni corrected). 21 compounds were exclusive to the DVB-CAR-PDMS SPME fibre and one compound was exclusive to the CAR-PDMS SPME fibre. A list of compounds exclusive to each fibre is in Supplementary Table 5. One compound (benzaldehyde) was significantly greater in abundance in samples exposed to the DVB-CAR-PDMS SPME fibre than the CAR-PDMS SPME fibre ($p < 0.05$, ANOVA, Bonferroni corrected). A chromatogram overlay was generated for the DVB-CAR-PDMS and CAR-PDMS fibres for one of the horses (H2) and is shown in Fig. 2.

3.3 Headspace vial volume

The mean (±SD) number of VOCs was 83 (±9.6) and 78 (±9.3) in 10 and 20 ml vials, respectively. Differences in mean VOC numbers between 10 and 20 ml vials were

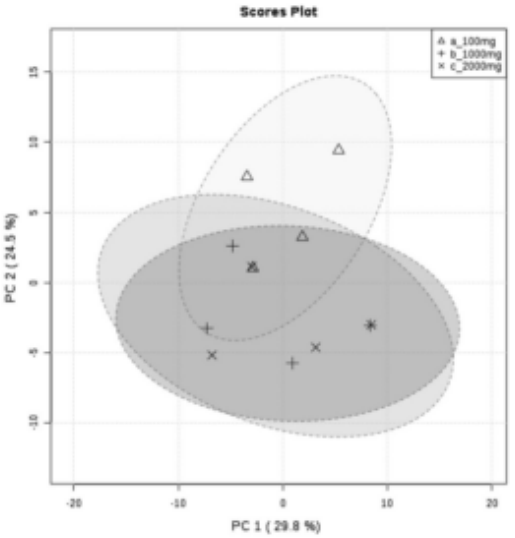


Fig. 1 A PCA of the VOC profiles of 100, 1000 and 2000 mg of horse faeces analysed by HS-SPME-GCMS. Grouped samples are contained within a 95% confidence interval

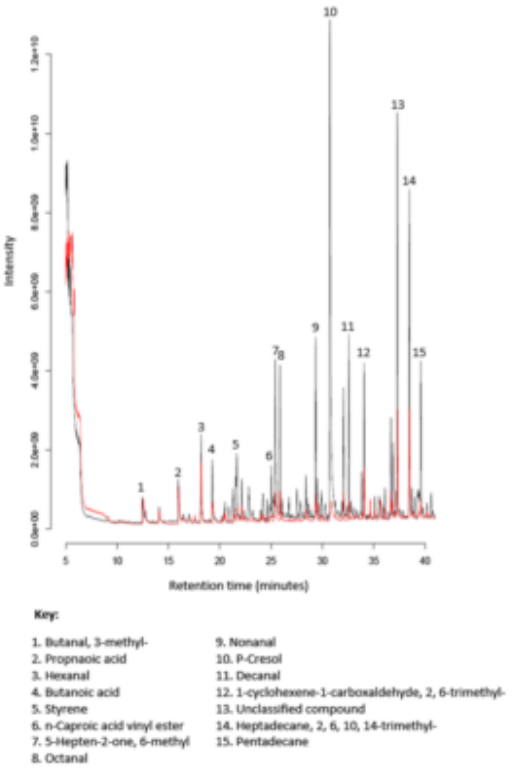


Fig. 2 An overlay of chromatograms generated from the HS-SPME-GCMS analysis of faeces of horse 2 (H2). The black trace represents the DVB-CAR-PDMS fibre and the red is CAR-PDMS

not significantly different from each other ($p = 0.48$, t test, Bonferroni corrected). Two compounds (propanoic acid, 2-methyl- and acetic acid, methyl ester) were exclusive to 20 ml vials and four compounds (propanoic acid, 2-methyl-, methyl ester, 2-hexanone, propanoic acid, 2-methyl-, propyl ester, 1-nonanol) were detected from 10 ml vials only. There was no significant difference in the abundance of any compounds shared between 10 and 20 ml vials ($p > 0.05$, t test, Bonferroni corrected).

3.4 Technical replicates

The CV was performed on the peak areas of VOCs shared across nine technical replicates of 1000 mg of faeces for each horse. The CV for each VOC peak area and the mean numbers of VOCs detected in each horse are in Supplementary Table 6. To summarise the CV for VOC peak area from H1 ranged between 0.7 and 8.6, 1–11% for H2, 0.9–8.3% for H3 and 0.7 to 9.6% for H4. A cluster analysis was performed based on the VOC profiles of all technical replicates and is shown in Fig. 3.

3.5 Time and temperature of storage

The mean number of VOCs identified in the fresh faecal sample, storage after 1 week, 6 months and 12 months at − 20° and − 80° are shown in Table 2.The percentages of VOCs shared between fresh and stored samples are also shown in Table 2. A PCA of the VOC profiles and a stack plot of the chemical classes of compounds found in each condition are shown in Fig. 4. A list of scores for PC1 and PC2 is supplied in Supplementary Table 7.

4 Discussion

4.1 Sample mass

A mass of 100 mg of faeces produced fewer VOCs than 1000 or 2000 mg. Masses of 1000 and 2000 mg showed very little variation between them and therefore 1000 mg may be considered an optimal mass for HS-SPME-GCMS of equine faeces. Masses greater than 2000 mg were not investigated as little difference was seen between 1000 and 2000 mg. It is likely that saturation of the SPME fibre or over-loading of the instrument has started to occur (Ng et al. 2012). The PCA plot (Fig. 1) accounts for 54.3% of the variance in the data set. Toluene, butanoic acid, 3-methyl-, ethyl ester and butanoic acid, 2-methylbutyl ester were among the VOCs most responsible for the variation seen in PC1. For PC2 styrene, butanoic acid, ethyl ester, D-limonene were VOCs most responsible for variation. A number VOCs with high scoring PCs were also found to be exclusive to 1000 and 2000 mg samples (Supplementary Table 2).

4.2 SPME fibre type

A greater number of VOCs was obtained when using DVB-CAR-PDMS fibre, rather than CAR-PDMS alone. The use of multiple fibre coatings increases the diversity of VOCs obtained (Dixon et al. 2011; Reade et al. 2014). The choice of fibre coating is an important factor affecting SPME as there is not a single fibre coating suited to all analytes. A DVB coating is mainly mesoporous, has a trapping range of C6-C15 and because of these properties is more suited to extracting medium and high molecular weight compounds (Mani 1999; Gianelli et al. 2002). Whereas a CAR coating is microporous, has a trapping range of C2–C12 and is more suited to low molecular weight compounds (Mani 1999; Gianelli et al. 2002). However, a more marked variation in VOC diversity between CAR-PDMS and

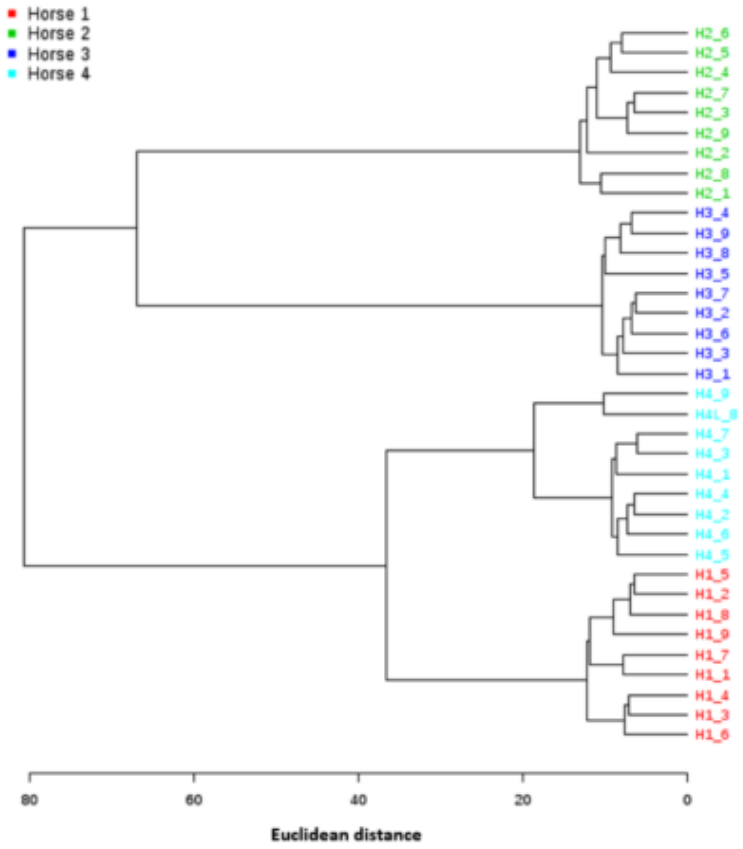


Fig. 3 A dendrogram constructed using the elucidean distance and ward clustering algorithm. Technical replicates of 1000 mg of faeces from four horses, analysed by HS-SPME-GCMS. VOCs were extracted using a DVB-CAR-PDMS SPME fibre

DVB-CAR-PDMS was observed in the present work than in human faeces (Couch et al. 2013). Therefore, the use of DVB and CAR fibre coatings, combined with PDMS, appears to yield the best results for equine faeces: this is supported by the findings of Bianchi et al. when studying short chain fatty acids from an in vitro colonic fermentation model (Bianchi et al. 2011). Therefore, a combination of the two coatings results in a net increase in diversity of compounds and offers a wider range of extraction. Comparisons made between SPME and other extraction techniques including direct thermal desorption (Cavalli et al. 2003) and purge and trap (Povolo and Contarini 2003) have shown to yield different VOCs. Alternatives to SPME could yield a greater library of VOCs and should be considered for future method development.

4.3 Headspace vial volume

The volume of the vial headspace containing 1000 mg of faeces did not have an effect on the number of VOCs obtained from the headspace of faeces. Furthermore, neither the 10 ml nor 20 ml vial demonstrated a clear advantage over the other in terms of detecting VOCs at a higher abundance. These findings agree with work by others that an increase in headspace volume does not have an impact on analyte detection (Cho et al. 2003). However it may depend on sample matrix, as found in murine faeces and human faeces (Reade et al. 2014). A smaller headspace volume resulted in a higher yield of VOCs for murine, whereas for human faeces (and horse in the present work) an alteration to the headspace volume had little impact. The SPME theory suggests that decreasing the volume of headspace increases the chance of compounds

Table 2 A table of the mean number of VOCs and percentages of VOCs shared between an equine faecal sample analysed by HS-SPME-GCMS 2 h post collection, after storage for 1 week at − 20 and − 80 °C and after 6 months at − 20 and − 80 °C

Storage conditions (°C)	Mean (± SD) numbers of VOCs
Fresh	79 (± 1.5)
1 week at − 20	99 (± 1.5)
1 week at − 80	82 (± 1.5)
6 months at − 20	95 (± 3.6)
6 months at − 80	80 (± 0.6)
12 months at − 20	93 (± 2.5)
12 months at − 80	80 (± 2.6)
Storage conditions (°C)	**Percentage (%) of VOCs shared**
Fresh and 1 week at − 20	79
Fresh and 6 months at − 20	65
Fresh and 12 months at − 20	64
Fresh and 1 week at − 80	91
Fresh and 6 months at − 80	84
Fresh and 12 months at − 80	84

being detected. A vial volume increase from 10 to 20 ml may not have provided a sufficient increase in headspace for a difference in VOCs to be seen in the present work. Greater numbers of VOCs may be extracted from samples with an increased exposure time or temperature (Reade et al. 2014). However an exposure time of 20 min and extraction temperature of 60 °C were optimal from those studied for both human and murine samples in our lab and therefore were not investigated here (Reade et al. 2014). An exposure time of longer than 20 min may also be impractical for large-scale studies requiring high-throughput analysis. Masses greater than 1000 mg were not investigated in 20 ml vials as significant differences between VOC numbers and abundances between 1000 and 2000 mg were not seen in 10 ml vials and it was thought that the SPME fibre may have reached its limit of absorbance at 1000 mg (Sect. 4.1). It can be concluded that both 10 and 20 ml vials are suitable for SPME-GC-MS analysis of 1000 mg horse faeces with little impact on the presence or abundance of VOCs obtained.

4.4 Technical replicates

Technical replicates of 1000 mg of faeces clustered closely according to the horse sampled and all samples fell within a 95% confidence interval (Supplementary Fig. 1).Within each horse the CV of shared VOC peak areas were below 11%, indicating a good reproducibility of the method.

4.5 Time and temperature of storage

Metabolic profiling of faeces using GC-MS analysis has been widely reported in numerous species. The processing of fresh samples is not always practical and therefore methods of preserving samples are necessary. However, the effect of storage length and temperature on faecal VOCs has received little attention (Deda et al. 2015). In the present work a higher number of compounds were found in samples stored at − 20 °C than fresh or − 80 °C. An increase in compounds after 1 month of freezing (− 80 °C) has been observed by others (Li et al. 2011). Chemical classification of VOCs revealed a large proportion of compounds in − 20 °C stored samples were esters. Furthermore VOCs accounting for the most variation in PC1 were largely esters (Supplementary Table 7). Fewer esters were seen in fresh samples and those stored at − 80 °C, indicating that the formation of esters is specific to − 20 °C storage. From this work it can only be speculated as to why esters were specific to storage at − 20 °C, possibly because at − 20 °C the sample may take longer to freeze and therefore esterification of acids continues compared to a sample frozen at − 80 °C. This phenomenon may be individual pony or storage vessel specific and requires further investigation. In human faeces it was observed that after 24 h of storage at − 20 °C there was no difference in VOC profile between that and a fresh sample (Gratton et al. 2016). Species differences are evident in preparation steps for metabolomics analysis (Reade et al. 2014). It is likely these differences are because of the differing VOC profiles between species, largely attributed to diet and intestinal microbiota. Unique storage conditions may be required for different types of sample. It was interesting to note that the percentage of VOCs shared between the fresh sample and samples stored at − 80 °C for 6 and 12 months decreased. An example includes benzene, (1-methylethyl)- which was present in a fresh sample and a sample stored at − 80 °C for 1 week, but was not present in any − 80 °C samples after 6 or 12 months. Whereas 1-octen-3-ol was not present in the fresh sample but was detected in − 80 °C samples after storage of 1 week, 6 months and 12 months. From this work it is difficult to determine why VOCs were lost or gained after storage as we still know very little of the impact of freezing on VOCs (Berkhout et al. 2016). We speculate the loss or gain of VOCs may be attributed to the material a sample is stored in (Mochalski et al. 2009) or the effect of freeze-thawing on microorganisms present in faeces (Achá et al. 2005). A limitation of this investigation is that a faecal sample was obtained from one animal. However, to the authors' knowledge this is the first work attempting to address the effect of long-term storage at varying temperatures on equine faecal VOCs and has highlighted the importance of these factors in experimental design. Based on this work a sample should be stored at − 80 °C and analysed

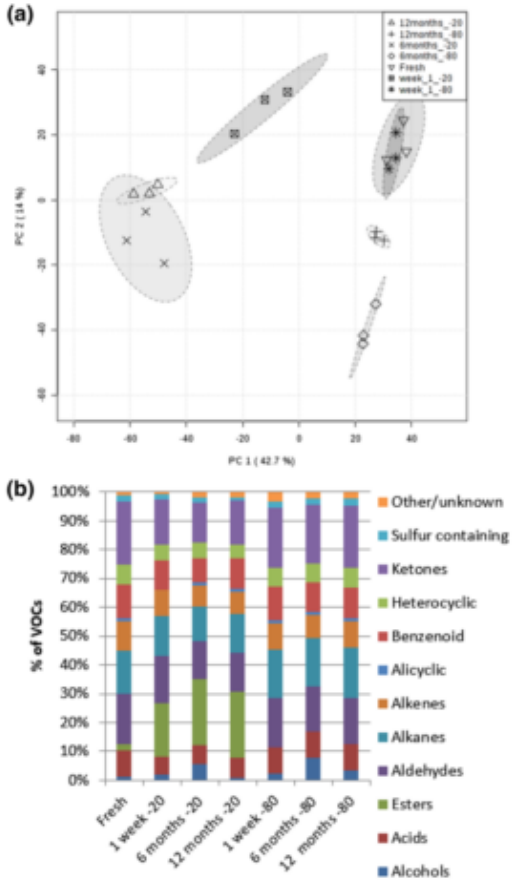


Fig. 4 **a** A PCA of the VOC profiles of an equine faecal sample analysed by HS-SPME-GCMS 2 h post collection, after storage for 1 week, 6 and 12 months at − 20 and − 80 °C. 3 **b** A stack plot of the chemical classes of compounds found in each storage condition

within 1 week to most resemble a fresh sample. Further work is necessary to confirm these findings in larger sample sizes and in other species.

5 Overall discussion

An optimal method for the preparation of samples for metabolomics is essential for achieving accurate and reproducible results. The preparation method for metabolic profiling differs between techniques selected, sample matrix and species. This work is the first of its kind working towards a robust method to extract VOCs from equine faeces using HS-SPME-GCMS. It was found that aliquots of 1000 mg in 10 or 20 ml SPME headspace vials were optimal. Volatile organic compounds should ideally be extracted using a DVB-CAR-PDMS SPME fibre. Faecal samples for VOC analysis to be stored at − 80 °C up to 12 months, where this is not possible all samples should be stored for the same length of time at the same temperature.

One of the main limitations of this work are the low number of replicates, particularly for the investigation of the storage of samples. A larger sample size may be better equipped to account for the influence of individual differences including breeds of horse, diet, age, effect of anthelmintic treatment etc. which may influence the faecal metabolome of the horse and hence the optimal method required for extraction of VOCs. However very few of these factors and their impact on the equine faecal VOC metabolome have been studied. The large gaps in measurement variables chosen here act only as a starting point and more precise intervals should feature in future work. Interactions between experimental factors were not investigated here. The reasons for this are that it was likely that saturation of the SPME fibre or over-loading of the instrument had started to occur at 1000 mg and there was no difference when this mass of sample was placed in a larger vial. Furthermore the CAR-PDMS fibre did not produce any advantage over 1000 mg in a 10 ml vial therefore we did not investigate any further sample masses with this fibre coating. A number of quality control techniques were applied in this work including the regular testing of laboratory air during analysis and blanks were run between samples to prevent carry-over. Samples were analysed in triplicate and the order of running samples was computer randomised. However, techniques to account for systematic variation e.g. the use of internal standards or pooled samples, were not explored here. As reviewed by (Dudzik et al. 2018) it is very important that robust quality control techniques are made available and should be explored in future method development work for extracting VOCs from equine faeces.

6 Conclusions

This work is a starting point for working towards a standardised method will allow comparisons to be made between future equine faecal GCMS based studies. Future work should involve larger sample sizes, consider individual characteristics and explore more precise measurement variables and quality control techniques.

Compliance with ethical standards

Conflict of interest RH, DA and CP declare that they have no conflict of interest.

Research involving animals Ethical approval was granted by the University of Liverpool ethics committee (ref: VREC279). All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

References

Achá, S. J., Kühn, I., Mbazima, G., Colque-Navarro, P., & Möllby, R. (2005). Changes of viability and composition of the Escherichia coli flora in faecal samples during long time storage. *Journal of Microbiological Methods, 63*, 229–238.

Aggio, R., Villas-Bôas, S. G., & Ruggiero, K. (2011). Metab: An R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics, 27*, 2316–2318.

Amann, A., Costello, B. D. L., Miekisch, W., Schubert, J., Buszewski, B., Pleil, J., et al. (2014). The human volatilome: Volatile organic compounds (VOCs) in exhaled breath, skin emanations,urine, feces and saliva. *Journal of Breath Research, 8*, 34001.

Arthur, C. L., & Pawliszyn, J. (1990). Solid phase microextraction with thermal desorption using fused silica optical fibers. *Analytical Chemistry, 62*, 2145–2148.

Bergman, E. N. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiological Reviews, 70*, 567–590.

Berkhout, D., Benninga, M., van Stein, R., Brinkman, P., Niemarkt, H., de Boer, N., et al. (2016). Effects of sampling conditions and environmental factors on fecal volatile organic compound analysis by an electronic nose device. *Sensors, 16*, 1967.

Bianchi, F., Dall'Asta, M., Del Rio, D., Mangia, A., Musci, M., & Scazzina, F. (2011). Development of a headspace solid-phase microextraction gas chromatography–mass spectrometric method for the determination of short-chain fatty acids from intestinal fermentation. *Food Chemistry, 129*, 200–205.

Cavalli, J. F., Fernandez, X., Lizzani-Cuvelier, L., & Loiseau, A. M. (2003). Comparison of static headspace, headspace solid phase microextraction, headspace sorptive extraction, and direct thermal desorption techniques on chemical composition of French olive oils. *Journal of Agricultural and Food Chemistry, 51*, 7709–7716.

Cho, D. H., Kong, S. H., & Oh, S. G. (2003). Analysis of trihalomethanes in drinking water using headspace-SPME technique with gas chromatography. *Water Research, 37*, 402–408.

Couch, R. D., Navarro, K., Sikaroodi, M., Gillevet, P., Forsyth, C. B., Mutlu, E., et al. (2013). The approach to sample acquisition and its impact on the derived human fecal microbiome and VOC metabolome. *PloS one, 8*, e81163.

Daly, K., Proudman, C. J., Duncan, S. H., Flint, H. J., Dyer, J., & Shirazi-Beechey, S. P. (2012). Alterations in microbiota and fermentation products in equine large intestine in response to dietary variation and intestinal disease. *The British Journal of Nutrition, 107*, 989–995.

Deda, O., Gika, H. G., Wilson, I., & Theodoridis, G. A. (2015). An overview of fecal sample preparation for global metabolic profiling. *Journal of Pharmaceutical and Biomedical Analysis, 113*, 137–150.

Dixon, E., Clubb, C., Pittman, S., Ammann, L., Rasheed, Z., Kazmi, N., et al. (2011). Solid-phase microextraction and human fecal VOC metabolome. *PloS one, 6*, e18471.

Dudzik, D., Barbas-Bernardos, C., García, A., & Barbas, C. (2018). Quality assurance procedures for mass spectrometry untargeted metabolomics: A review. *Journal of Pharmaceutical and Biomedical Analysis, 147*, 149–173.

Garner, C. E., Smith, S., de Lacy Costello, B., White, P., Spencer, R., Probert, C. S. J., et al. (2007). Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *FASEB Journal, 21*, 1675–1688.

Garner, C. E., Smith, S., Elviss, N. C., Humphrey, T. J., White, P., Ratcliffe, N. M., et al. (2008). Identification of campylobacter infection in chickens from volatile faecal emissions. *Biomarkers, 13*, 413–421.

Gianelli, M. P., Flores, M., & Toldra, F. (2002). Optimisation of solid phase microextraction (SPME) for the analysis of volatile compounds in dry-cured ham. *Journal of the Science of Food and Agriculture, 82*, 1703–1709.

Gratton, J., Phetcharaburanin, J., Mullish, B. H., Williams, H. R. T., Thursz, M., Nicholson, J. K., et al. (2016). Optimized sample handling strategy for metabolic profiling of human feces. *Analytical Chemistry, 88*, 4661–4668.

Ishizaka, S., Matsuda, A., Amagai, Y., Oida, K., Jang, H., Ueda, Y., et al. (2014). Oral administration of fermented probiotics improves the condition of feces in adult horses. *Journal of Equine Science, 25*, 65–72.

Li, J. V., Saric, J., Wang, Y., Keiser, J., Utzinger, J., & Holmes, E. (2011). Chemometric analysis of biofluids from mice experimentally infected with Schistosoma mansoni. *Parasites & Vectors, 4*, 179.

Mani, V. (1999). Properties of commercial SPME fibre coatings. In J. Pawliszyn (Ed.), *Applications of solid phase microextraction* (pp. 57–108). Cambridge: The Royal Society of Chemistry.

Marcobal, A., Kashyap, P. C., Nelson, T. A., Aronov, P. A., Donia, M. S., Spormann, A., et al. (2013). A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME Journal, 7*, 1933–1943.

Milinovich, G. J., Trott, D. J., Burrell, P. C., Croser, E. L., Al Jassim, R. A., Morton, J. M., et al. (2007). Fluorescence in situ hybridization analysis of hindgut bacteria associated with the development of equine laminitis. *Environmental Microbiology, 9*, 2090–2100.

Mochalski, P., Wzorek, B., Śliwka, I., & Amann, A. (2009). Suitability of different polymer bags for storage of volatile sulphur compounds relevant to breath analysis. *Journal of Chromatography B, 877*, 189–196.

Povolo, M., & Contarini, G. (2003). C omparison of solid-phase microextraction and purge-and-trap methods for the analysis of the volatile fraction of butter. *Journal of Chromatography A, 985*, 117–125.

Proudman, C. J., Hunter, J. O., Darby, aC., Escalona, E. E., Batty, C., & Turner, C. (2014). Characterisation of the faecal metabolome and microbiome of Thoroughbred racehorses. *Equine Veterinary Journal, 47*, 580–586.

Reade, S., Mayor, A., Aggio, R., Khalid, T., Pritchard, D., Ewer, A., et al. (2014). Optimisation of sample preparation for direct SPME-GC-MS analysis of murine and human faecal volatile organic compounds for metabolomic studies. *Journal of Analytical & Bioanalytical Techniques, 5*, 184.

Saric, J., Wang, Y., Li, J., Coen, M., Utzinger, J., Marchesi, J. R., et al. (2008). Species variation in the fecal metabolome gives insight into differential gastrointestinal function. *Journal of Proteome Research, 7*, 352–360.

Stavert, J. R., Drayton, B. A., Beggs, J. R., & Gaskett, A. C. (2014). The volatile organic compounds of introduced and native dung

and carrion and their role in dung beetle foraging behaviour. *Ecological Entomology, 39*, 556–565.

Ng, J. S. Y., Ryan, U., Trengove, R., & Maker, G. (2012). Development of an untargeted metabolomics method for the anlysis of human faecal samples using *Cryptosporidium*-infected samples. *Molecular & Biochemical Parasitology, 185*, 145–150.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics, 3*, 211–221.

Turner, C., Batty, C., Escalona, E., Hunter, J., & Proudman, C. (2013). The use of SIFT-MS in profiling the faecal volatile metabolome in horses with colic:A pilot study. *Current Analytical Chemistry, 9*, 614–621.

Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D., & Wishart, D. S. (2012). MetaboAnalyst 2.0-a comprehensive server for metabolomic data analysis. *Nucleic Acids Research, 40*, 127–133.

PeerJ

A longitudinal study of the faecal microbiome and metabolome of periparturient mares

Shebl E. Salem[1,2], Rachael Hough[3], Chris Probert[3], Thomas W. Maddox[4], Philipp Antczak[5], Julian M. Ketley[6], Nicola J. Williams[1], Sarah J. Stoneham[7] and Debra C. Archer[1]

[1] Department of Epidemiology and Population Health, Institute of Infection and Global Health, University of Liverpool, Leahurst campus, Wirral, UK
[2] Department of Surgery, Faculty of Veterinary Medicine, Zagazig University, Zagazig, Al Sharquiya, Egypt
[3] Department of Cellular and Molecular Physiology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK
[4] Department of Musculoskeletal Biology, Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, UK
[5] Computational Biology Facility, Institute of Integrative Biology, University of Liverpool, Liverpool, UK
[6] Department of Genetics and Genome Biology, College of Life Sciences, University of Leicester, Leicester, UK
[7] TopSpec Equine Ltd, Middle Park Farm, North Yorkshire, UK

ABSTRACT

Background: Periparturient mares are at increased risk of colic including large colon volvulus, which has a high mortality rate. Alterations in colonic microbiota related to either physiological or management changes, or both, that occur at this time have been suggested as potential causes for increased colic risk in this population of horses. Although the effect of management changes on the horse faecal microbiota has been investigated, limited work has been conducted to investigate changes in faecal microbiota structure and function in the periparturient period.
The objectives of the current study were to investigate temporal stability of the faecal microbiota and volatile organic compounds (VOCs) of the faecal metabolome in periparturient mares.
Methods: Faecal samples were collected weekly from five pregnant mares from 3 weeks pre-foaling to 7 weeks post-foaling. The microbiome data was generated by PCR amplification and sequencing of the V1–V2 regions of the bacterial 16S rRNA genes, while the VOC profile was characterised using headspace solid phase microextraction gas chromatography mass spectrometry.
Results: The mare faecal microbiota was relatively stable over the periparturient period and most variation was associated with individual mares. A small number of operational taxonomic units were found to be significantly differentially abundant between samples collected before and after foaling. A total of 98 VOCs were identified. The total number of VOCs did not vary significantly between individual mares, weeks of sample collection and feeds available to the mares. Three VOCs (decane, 2-pentylfuran, and oct-2-ene) showed significant increase overtime on linear mixed effects modelling analysis. These results suggest that the mare faecal microbiota is structurally and functionally stable during the periparturient period.

Submitted 15 January 2019
Accepted 26 February 2019
Published 3 April 2019

Corresponding authors
Shebl E. Salem, ssalem@rvc.ac.uk
Debra C. Archer, darcher@liverpool.ac.uk

Academic editor
Joseph Gillespie

Additional Information and Declarations can be found on page 12

DOI 10.7717/peerj.6687

How to cite this article Salem SE, Hough R, Probert C, Maddox TW, Antczak P, Ketley JM, Williams NJ, Stoneham SJ, Archer DC. 2019. A longitudinal study of the faecal microbiome and metabolome of periparturient mares. **PeerJ 7:e6687 DOI 10.7717/peerj.6687**

compared (*Weese et al., 2015*). However, the risk period for postpartum colic in broodmares appears to be greatest in the first 90 days post-foaling and the latter study did not investigate this longer period.

Volatile organic compounds (VOCs) are the products of metabolism of the microbiota and the host (mare). These compounds provide information about the functional microbiota (i.e. what the bacteria produce), which may be more important than identification of bacteria alone. In humans and animals, altered VOC profiles were shown to be non-invasive indicators of gastrointestinal disease (*Garner et al., 2007*; *Leng et al., 2018*). Therefore, faecal VOC analysis may eventually provide a monitoring tool for types of equine colic that may arise from gut dysbiosis. The aim of the current study was to characterise the faecal microbiota and the faecal volatile metabolome of periparturient mares from 3 weeks pre-foaling to 7 weeks post-foaling, and to explore temporal stability of structure of the microbial community and the VOCs it produces.

MATERIALS AND METHODS

Mares

Five healthy pregnant mares from the same farm were recruited onto the study approximately 1 month prior to their foaling due dates. The mares had no history of colic or other medical conditions during pregnancy. The demographics of the mares are summarised in Table S1. The mares' diet was supplemented with a feed balancer (Opti-Care Balancer; Gain Equine Nutrition, Durham, UK) throughout the study and they had free access to water from automatic watering devices. Pre-foaling and during the first few weeks post-foaling the mares were managed on grass paddocks during the day and stabled at night in separate foaling boxes where they were fed hay. Mares were then turned out in groups in larger grass paddocks with free access to lush grass until the end of the study. The timing of introduction of each of these nutritional management practices varied between mares and details are given in Fig. S1. Moxidectin (Equest, Zoetis, Surrey, UK) was administered to all mares approximately 2 months prior to collection of the first set of samples, which was part of the normal management routine of the mares. The study was approved by The University of Liverpool Veterinary Research Ethics Committee (VREC207) and the manager of the farm consented to participate.

Sample collection

A total of 11 samples (approximately 200 g each) were collected weekly from faeces passed by individual mares immediately following observed defaecation (samples labelled T-3 to T7). Samples were placed in plastic sealable bags and were stored in a refrigerator (4 °C) at the farm temporarily until all samples had been collected from mares scheduled for sampling on that day. Samples were then transferred to a −80 °C freezer where they were stored until processing. All samples were frozen within 5 h of collection except samples collected during the second stage labour (T0) where the yard staff collected these samples and refrigerated them until the next visit by the principal investigator. The sampling schedule and types of feed at each sampling occasion are depicted in Fig. S1.

DNA extraction and generation of sequence data

A commercial kit (QIAamp DNA Stool Mini Kit; QIAGEN, Manchester, UK) was used to extract DNA from samples followed by PCR amplification of the V1–V2 hypervariable regions of the bacterial 16S ribosomal ribonucleic acid (rRNA) gene to create amplicon libraries for sequencing using the Ion Torrent Personal Genome Machine system. Details of sample preparation and generation of sequence data are described previously (*Salem et al., 2018*).

Faecal volatile organic compound profiling

The VOC profile of faeces was determined using headspace solid phase microextraction gas chromatography mass spectrometry (HS-SPME-GCMS) using the methods described by *Hough, Archer & Probert (2018)*.

Data analysis

Microbiome data analysis

The data generated from sequencing of the 16S rRNA gene amplicon libraries were processed using the QIIME pipeline (version 1.8.0; http://qiime.org/) (*Caporaso et al., 2010b*). Sequences from different samples were demultiplexed according to their barcode sequences and chimeric sequences were identified using the UCHIME algorithm (*Edgar et al., 2011*) and were filtered from the data. Sequences were then clustered open-reference into operational taxonomic units (OTUs) at 97% identity threshold using USEARCH (version 6.1.544) (*Edgar, 2010*). A representative sequence for each OTU cluster was aligned to the Greengenes core set (version 13.8) (*DeSantis et al., 2006*) using PyNAST (*Caporaso et al., 2010a*), filtered to remove gaps and hypervariable regions using the Lane mask before creating an approximately-maximum-likelihood phylogenetic tree using FastTree (*Price, Dehal & Arkin, 2010*). Taxonomic assignment of OTU representatives was performed using the ribosomal database project classifier (version 2.2) (*Wang et al., 2007*) informed with the Greengenes reference database at a 50% confidence limit.

Statistical analyses were performed using R software environment (version 3.2.2) (*R Core Team, 2014*) with the following add-on statistical packages: 'phyloseq' (*McMurdie & Holmes, 2013*), 'vegan' (*Oksanen et al., 2015*), 'ggplot2' (*Wickham, 2009*), 'nlme' (*Pinheiro et al., 2015*), and 'cluster' (*Maechler et al., 2016*). Apart from alpha diversity analysis, the OTU table was further filtered to remove low-abundance OTUs (OTUs present in <5% of samples or represented by <20 reads) and normalised by rarefying to account for unequal sequencing effort between samples (*Weiss et al., 2015*).

The data were clustered hierarchically based on the average linkage agglomerative clustering method (UPGMA) following calculation of a Bray–Curtis dissimilarity matrix. The trees were visualised using publicly available software (FigTree version 1.4.2). Principal coordinate analysis (PCoA) was performed on a Bray–Curtis dissimilarity matrix created from the OTU table. The amount of variation in the data that could be explained by either the time of sample collection relative to foaling, feed or the individual mares was estimated using permutational multivariate analysis of variance (PERMANOVA)

following calculation of a Bray–Curtis dissimilarity matrix from the data using the 'vegan::adonis' function in R.

The statistical methods and R script used for investigating changes in faecal microbiota diversity and stability over time in the current study were adapted from publicly available R codes (*DiGiulio et al., 2015a*). Alpha diversity analysis involved calculation of Chao1 index (*Chao, 1984*) for species richness (number of different species within a community); and Shannon (*Shannon, 1948*) diversity index for population diversity (a measure of species richness and similarity of species abundance within a community). The pattern of change of calculated measures over time relative to foaling were evaluated using linear mixed-effects modelling (LME). Random intercept and slope LME models were fitted where mares were included as a random effect and time relative to foaling (in weeks) was included as a fixed effects term in the model.

Beta diversity analysis involved calculation of weighted-UniFrac (*Lozupone et al., 2007*) and Bray–Curtis (*Bray & Curtis, 1957*) dissimilarity metrics. Distances between consecutive samples of the community (distances between consecutive sampling time points) within each of the mares were calculated and were used as a measure of stability of the community over time relative to foaling. The calculated distances were modelled using random intercept and slope LME models where mares were treated as a random effect variable and where time was the fixed portion of the model.

Samples collected during a 3-week period before and after foaling, were compared for differentially abundant OTUs using negative binomial models. The models were fitted using the DESeq2::DESeq function in R. Prior to differential abundance analysis, the OTU table was further filtered to exclude OTUs present in <25% of the samples. *p*-values were adjusted for multiple testing using the false discovery rate (FDR) method (*Benjamini & Hochberg, 1995*). OTUs the adjusted *p*values of which were <0.1 were considered significant. Results from this model were presented in a dot plot.

Metabolome data analysis

Metabolome data were processed using Automated Mass Spectral Deconvolution System (AMDIS-version 2.71, 2012) and VOCs were putatively identified using the National Institute of Standards and Technology mass spectral library (version 2.0, 2011). Data were aligned using the R package Metab (*Aggio, Villas-Boas & Ruggiero, 2011*). All samples were analysed in triplicate, an average was taken of the technical replicates and taken forward for data analysis. To allow statistical comparison of relative compound abundance, any missing values present after taking an average were replaced with a half-minimum value of the data matrix. Clustering within the data was investigated using principal component analysis (PCA) and the mean number of VOCs identified were compared between individual mares, weeks of sample collection and the types of feeds using ANOVA followed by Tukey's HSD test. PERMANOVA was used to estimate the amount of variation in the data that could be explained by individual mares, week of sample collection and the type of feed. LME modelling was used to explore the pattern of change of individual VOCs over time. For each VOC a Random intercept model was fitted where mares were included as a random effect variable and time in weeks was

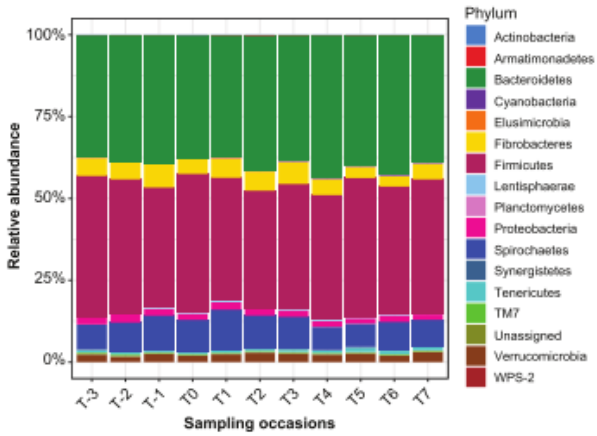


Figure 1 Relative abundance of bacterial phyla identified in the mare faecal microbiota. A bar plot of the relative abundance of different bacterial phyla identified in the data.
Full-size DOI: 10.7717/peerj.6687/fig-1

included as a fixed effects term in the model. VOCs with p-values of <0.1 following adjustment for multiple comparisons using the FDR method were considered significant. All analyses were performed using R (version 3.2.2).

RESULTS

Mare faecal microbiota

Sequencing of PCR-amplified 16S rRNA genes from 55 samples resulted in 1,648,876 quality non-chimeric sequences. Each sample had at least 17,392 reads, and there were an average of 29,980 reads per sample. The reads were clustered into 17,863 OTUs. Filtration of spurious OTUs reduced this count to 7,843 OTUs (40% of the original count). In this filtered OTU count table, 16 bacterial phyla were identified (Fig.1), the relative abundance of which at each sampling time point is presented in Table S2.

Cluster analysis revealed that the data were clustered by mares rather than by time relative to foaling. UPGMA trees built from a Bray–Curtis dissimilarity matrix is provided in Fig. S2. A similar pattern of clustering was also confirmed in a PCoA plot (Fig. 2). PERMANOVA analysis revealed that 3% ($R^2 = 0.03$, p-value $= 0.001$) of variation in the data could be explained by the variable time, 5% ($R^2 = 0.05$, p-value $= 0.001$) by the type of feed and 33% ($R^2 = 0.33$, p-value $= 0.001$) by individual mares.

Linear mixed-effects modelling of Chao1 and Shannon index diversity measures revealed general stability of alpha diversity over time. Prediction plots from these models are given in Figs. 3A and 3B. Similar results were also obtained when investigating dissimilarity (beta diversity) between consecutive sampling time points (Figs. 3C and 3D). These results suggest that the mare faecal microbiota was stable from 3 weeks pre-foaling to 7 weeks post-foaling. Only 81 OTUs were found to be significantly differentially

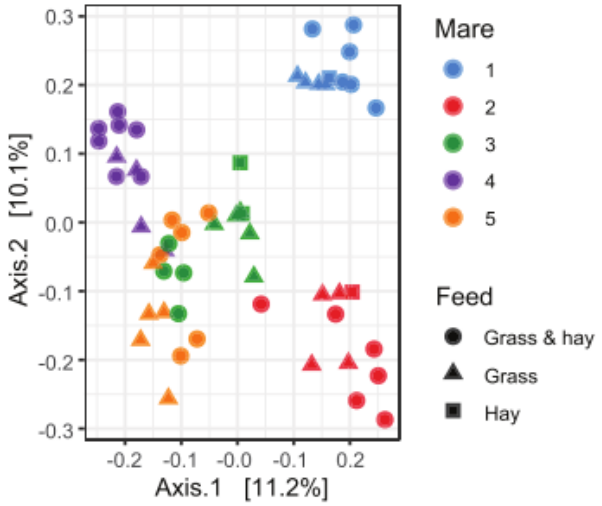


Figure 2 Principal coordinate analysis of the mare faecal microbiota. Ordination plot of the first two axes from the principal coordinate analysis (PCoA) of microbiome data. Data points are coloured by individual mares and shaped by the types of feed. Full-size DOI: 10.7717/peerj.6687/fig-2

abundant between samples collected during a 3-week period prior to foaling and a 3-week period post foaling (Fig. S3). These results also confirm greater stability of the mare faecal microbiota around the time of foaling. Of these OTUs, 54 were upregulated and 27 were downregulated post-foaling. Downregulated OTUs belonged mainly to the Ruminococcaceae family ($n = 39$) while the upregulated OTUs were represented mainly by members of Spirochaetaceae family ($n = 6$) (Table S3).

Mare faecal metabolome

A total of 98 VOCs were characterised in the faecal metabolome in all samples. The mean number of VOCs did not differ significantly between individual mares, weeks of sample collection or the types of feed (Figs. S4A–S4C). There was no clear clustering of the data on PCA analysis (Figs. 4 and 5). Similar to the microbiome data, PERMANOVA analysis showed that a small amount of variation could be explained by the time of sampling ($R^2 = 0.05$, $p = 0.01$) and type of feed ($R^2 = 0.06$, $p = 0.04$), while individual mares accounted for 12% of variation in the data ($R^2 = 0.12$, $p = 0.04$). LME modelling identified three VOCs which increased significantly (adjusted p-value < 0.1) over time during the study period (Fig. 6). These were decane, 2-pentylfuran and oct-2-ene.

DISCUSSION

The current study has characterised the faecal microbiota and the faecal metabolome of a group of pregnant mares during the periparturient period, at times when broodmares are at greatest risk of colic. We found that these microbial communities were structurally

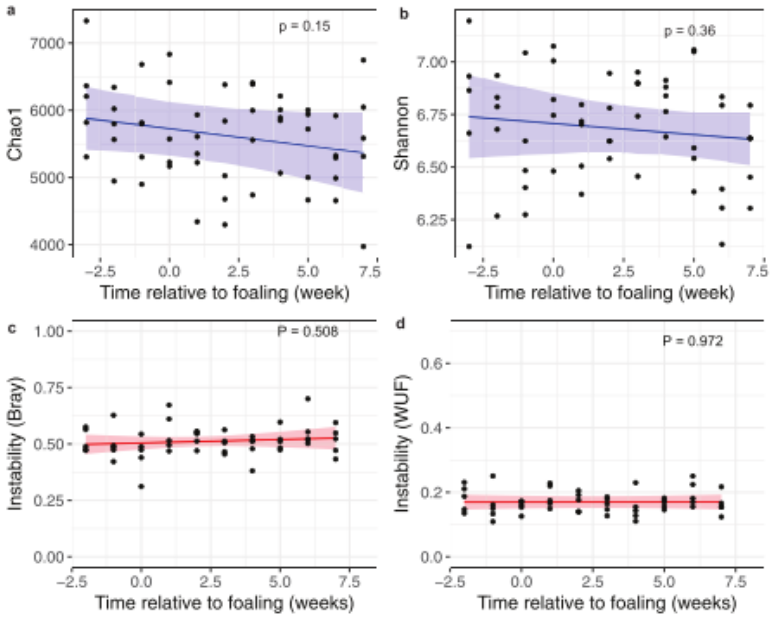


Figure 3 Changes in alpha and beta diversities over time. Prediction plots of (A) Chao1 and (B) Shannon diversity measures; and (C) a Bray–Curtis and (D) a Weighted Unifrac dissimilarity matrices against time relative to foaling. Blue and red lines are the regression lines from the linear mixed-effects models and cognate shading denotes the 95% confidence limits of the prediction. The models included time as a fixed effects variable and the mares as a random effect variable.

Full-size DOI: 10.7717/peerj.6687/fig-3

and functionally stable over the course of the study, and changes that were identified were largely associated with individual mares rather than being related to the time of sample collection. These findings are in agreement with previous studies (*Blackmore et al., 2013*; *Proudman et al., 2015*) and suggest that if changes in faecal microbiota are associated with postpartum colic, altered risk may be due to inherent differences between individual mares.

The findings of the current study are consistent with a recent human study that investigated the microbiota composition of four different body sites including the distal gut, vagina, saliva, and tooth/gums of pregnant women (*DiGiulio et al., 2015b*). The latter study found that these microbial communities were stable throughout pregnancy and post-delivery. Our findings are also consistent with those reported by *Weese et al. (2015)* who reported a minimal effect of foaling on the mare faecal microbiota. None of the mares included in the current study developed colic either pre- or postpartum and therefore the results reported here cannot be related directly to colic occurrence. However, our results suggest that it is unlikely that inherent changes in faecal microbiota

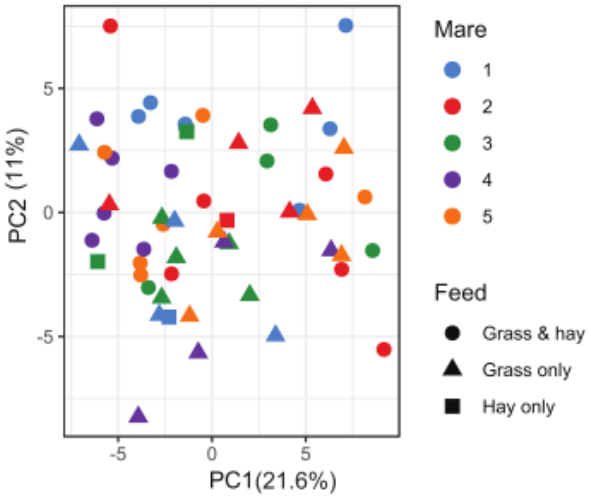


Figure 4 Principal component analysis of metabolome data. Ordination plot of the first two axes from the principal component analysis (PCA) of metabolome data. The plot shows that there is no clear clustering of the data by either the individual mares or the types feeds available to the mares. Data points are coloured by individual mares and shaped by the types of feeds.

Full-size DOI: 10.7717/peerj.6687/fig-4

of mares during the periparturient period are a major contributing factor in the development of periparturient colic. *Weese et al. (2015)* reported significant increase in relative abundance of the phylum Proteobacteria and reduction in relative abundance of the Firmicutes, Bacteroidetes, and Tenericutes phyla prior to the onset of postpartum colic. This observation requires further investigation to determine if such markers can identify mares at increased risk of colic and any interventions that may reduce this risk.

The level of alpha diversity of the mare faecal microbiota exhibited an overall linear stability with time relative to foaling. Alpha diversity is a measure of within-sample biodiversity and is often used to associate the perturbation of microbial communities with a disease process or changes in the physiological status of the host (*De Weerth et al., 2013*; *Elli, Colombo & Tagliabue, 2010*; *Turnbaugh et al., 2009*). Changes in beta diversity overtime were also minimal which together suggest that the mare faecal microbiota exhibited minimal change during the period under investigation. The significant increase in mean count of OTUs that belonged to the Ruminococcaceae family and the decrease in members of Spirochaetaceae family post-foaling could be attributed to changes in feed types where mares were predominantly fed on grass post-foaling. These findings concur with a previous study, which reported increase in relative abundance of members of Ruminococcaeae and decrease in relative abundance of members of Spirochaetaceae bacterial families in response to grass feeding (*Salem et al., 2018*).

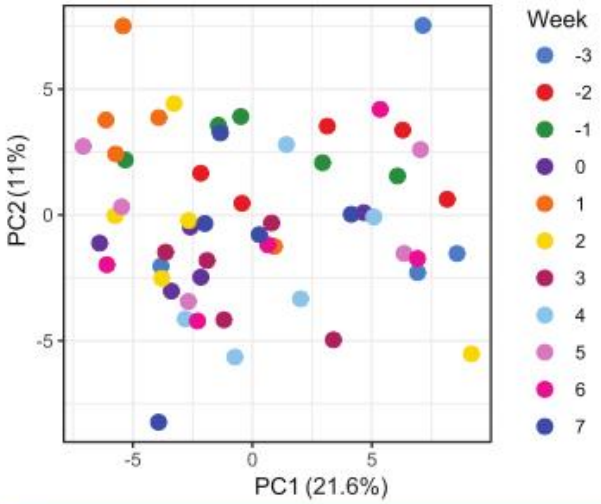


Figure 5 Principal component analysis of metabolome data. Ordination plot of the first two axes from the principal component analysis (PCA) of metabolome data. The plot shows that there is no clear clustering of the data by sampling time points. Data points are coloured by the time of sampling relative to foaling. Full-size DOI: 10.7717/peerj.6687/fig-5

Three compounds (decane, 2-pentylfuran, and oct-2-ene) altered significantly with time in the current study, while the remainder of the faecal VOC metabolome of the mares remained stable.

Decane, 2-pentylfuran, and oct-2-ene have been detected from the faecal headspace of ruminants and they may originate directly from food or as products of microbial fermentation (*Cai et al., 2006*; *Fischer et al., 2015*; *Garner et al., 2007*; *Schulz & Dickschat, 2007*). Therefore, the significant changes in these VOCs in the current study may reflect a change in feed type or mare feed intake over time. In general, the faecal VOC metabolome appeared to be more uniform among individual mares than the faecal microbiota, indicating a common functionality between bacteria. Overall, the microbiome and metabolome mirrored each other, in that they both remained stable, indicating that VOCs may be a cost-effective alternative to monitoring the faecal microbiome.

In the current study, the faecal microbiota was used as a proxy for the hindgut microbial populations. Studies that compared the microbiota composition of different regions of the horse gastrointestinal tract have found that faecal microbiota partially represent the microbial populations of the large intestine, particularly those of the distal colon (*Costa et al., 2015*; *Dougal et al., 2012*). Given that it is not possible to directly sample the colonic microbiota sequentially, faecal analysis is the best possible proxy measure of potential changes in the colonic microbial community. The effect of a new diet on the equine gut microbiota might not be evident until 4–6 days following dietary change (*Fernandes et al., 2014*; *Van Den Berg et al., 2013*). Ideally, mares maintained on one diet

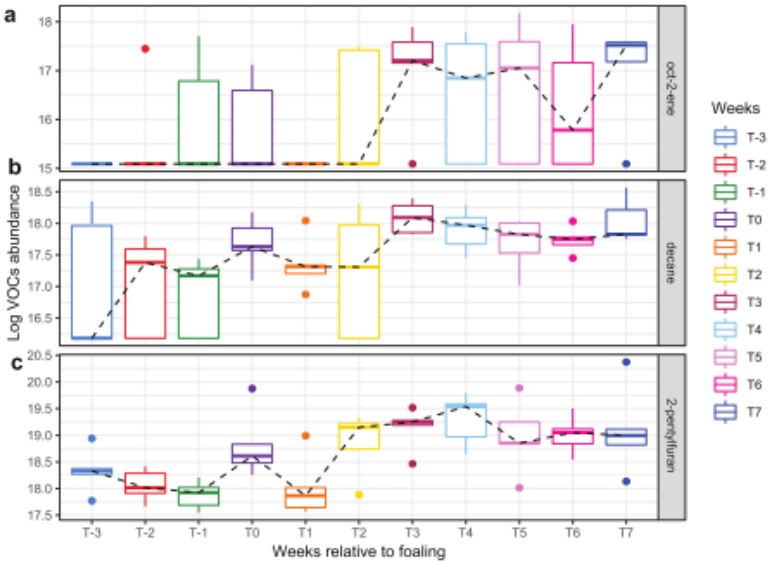


Figure 6 Boxplots of three VOCs that significantly increased over time during the study period. Boxplots of three VOCs ((A) oct-2-ene, (B) decane, (C) 2-pentylfuran) that significantly increased over time during the study period. Changes of VOC abundance over time was modelled using linear mixed-effects modelling with the mare included as a random effect variable and the time (in weeks) as a fixed effects variable in the model. Full-size DOI: 10.7717/peerj.6687/fig-6

could have been compared to groups of mares on differing diets. This was not possible in the current study nor does it reflect how broodmares are generally managed on many stud farms. Transition between these diets was abrupt, and it was not known exactly when the faecal microbiota changed in relation to this.

CONCLUSIONS

This is the first study to perform a detailed investigation of faecal microbiota and faecal metabolome of a group of mares during the periparturient period when they are at increased risk of colic. The study demonstrated that the majority of changes identified in the faecal microbiota and VOCs were mare-specific, and did not appear to be related to inherent physiological changes associated with foaling. The change in three VOCs post-parturition is interesting and may reflect a subtle change in the functional microbiome. Further studies are warranted to identify mares with altered metabolome in the periparturient period to determine whether there is a link to the risk of colic. VOCs may provide a cost-effective means of monitoring such mares with the ultimate aim of developing stable-side tests to identify and monitor mares at increased risk of colic.

ACKNOWLEDGEMENTS

The authors are grateful to the staff at the Cheshire Equine Clinic and the Harthill Stud Farm for their cooperation in sample collection. We would also like to thank Mr. A. Mayor (University of Liverpool) for writing some of the R functions used in the metabolome data analysis. We would like also to thank Dr. Adam Berg for advice on laboratory techniques and his contribution to initial bioinformatics analysis of the sequence data.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Shebl E Salem's PhD studies were funded by the Egyptian Ministry of Higher Education. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosure

The following grant information was disclosed by the authors:
The Egyptian Ministry of Higher Education.

Competing Interests

The authors declare that they have no competing interests. Sarah J. Stoneham was employed by the Cheshire Equine Clinic during the period of sample collection and she is currently employed as an equine nutritionist by TopSpec Equine Ltd. North Yorkshire, UK.

Author Contributions

- Shebl E. Salem conceived and designed the experiments, performed the experiments, analysed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Rachael Hough conceived and designed the experiments, performed the experiments, analysed the data, authored or reviewed drafts of the paper, approved the final draft.
- Chris Probert conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Thomas W. Maddox conceived and designed the experiments, approved the final draft.
- Philipp Antczak analysed the data, contributed reagents/materials/analysis tools, approved the final draft.
- Julian M. Ketley conceived and designed the experiments, contributed reagents/materials/analysis tools, approved the final draft.
- Nicola J. Williams conceived and designed the experiments, contributed reagents/materials/analysis tools, approved the final draft.
- Sarah J. Stoneham conceived and designed the experiments, approved the final draft.
- Debra C. Archer conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The University of Liverpool Veterinary Research Ethics Committee approved the study (VREC207).

Data Availability

The following information was supplied regarding data availability:

The raw sequence data, Qiime script and the R codes used for data analyses are available at figshare: Salem, Shebl (2019): A longitudinal study of the faecal microbiome and metabolome of periparturient mares. figshare. Fileset. DOI 10.6084/m9.figshare.7267343.v1.

Sequence data are available at the NIH Sequence Read Archive, accession number PRJNA523602.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.6687#supplemental-information.

REFERENCES

Aggio R, Villas-Boas SG, Ruggiero K. 2011. Metab: an R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics* **27(16)**:2316–2318 DOI 10.1093/bioinformatics/btr379.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57(1)**:289–300 DOI 10.1111/j.2517-6161.1995.tb02031.x.

Blackmore TM, Dugdale A, Argo CMG, Curtis G, Pinloche E, Harris PA, Worgan HJ, Girdwood SE, Dougal K, Newbold CJ, McEwan NR. 2013. Strong stability and host specific bacterial community in faeces of ponies. *PLOS ONE* **8(9)**:e75079 DOI 10.1371/journal.pone.0075079.

Bray JR, Curtis JT. 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs* **27(4)**:326–349 DOI 10.2307/1942268.

Cai L, Koziel JA, Davis J, Lo YC, Xin H. 2006. Characterization of volatile organic compounds and odors by in-vivo sampling of beef cattle rumen gas, by solid-phase microextraction, and gas chromatography–mass spectrometry–olfactometry. *Analytical and Bioanalytical Chemistry* **386(6)**:1791–1802 DOI 10.1007/s00216-006-0799-1.

Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26(2)**:266–267 DOI 10.1093/bioinformatics/btp636.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7(5)**:335–336 DOI 10.1038/nmeth.f.303.

Chao A. 1984. Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**:265–270.

Cohen ND, Gibbs PG, Woods AM. 1999. Dietary and other management factors associated with colic in horses. *Journal of the American Veterinary Medical Association* **215(1)**:53–60.

Cohen ND, Matejka PL, Honnas CM, Hooper RN. 1995. Case-control study of the association between various management factors and development of colic in horses. Texas equine colic study group. *Journal of the American Veterinary Medical Association* **206(5)**:667–673.

Cohen ND, Peloso JG. 1996. Risk factors for history of previous colic and for chronic, intermittent colic in a population of horses. *Journal of the American Veterinary Medical Association* **208(5)**:697–703.

Costa MC, Silva G, Ramos RV, Staempfli HR, Arroyo LG, Kim P, Weese JS. 2015. Characterization and comparison of the bacterial microbiota in different gastrointestinal tract compartments in horses. *Veterinary Journal* **205(1)**:74–80 DOI 10.1016/j.tvjl.2015.03.018.

Daly K, Proudman CJ, Duncan SH, Flint HJ, Dyer J, Shirazi-Beechey SP. 2012. Alterations in microbiota and fermentation products in equine large intestine in response to dietary variation and intestinal disease. *British Journal of Nutrition* **107(7)**:989–995 DOI 10.1017/S0007114511003825.

De Weerth C, Fuentes S, Puylaert P, De Vos WM. 2013. Intestinal microbiota of infants with colic: development and specific signatures. *Pediatrics* **131(2)**:e550–e558 DOI 10.1542/peds.2012-1449.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* **72(7)**:5069–5072 DOI 10.1128/AEM.03006-05.

DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Goltsman DS, Wong RJ, Shaw G, Stevenson DK, Holmes SP, Relman DA. 2015a. Temporal and spatial variation of the human microbiota during pregnancy. *Available at http://statweb.stanford.edu/~susan/papers/PNASRR.html* (accessed 19 May 2015).

DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Goltsman DS, Wong RJ, Shaw G, Stevenson DK, Holmes SP, Relman DA. 2015b. Temporal and spatial variation of the human microbiota during pregnancy. *Proceedings of the National Academy of Sciences of the United States of America* **112(35)**:11060–11065 DOI 10.1073/pnas.1502875112.

Dougal K, Harris PA, Edwards A, Pachebat JA, Blackmore TM, Worgan HJ, Newbold CJ. 2012. A comparison of the microbiome and the metabolome of different regions of the equine hindgut. *FEMS Microbiology Ecology* **82(3)**:642–652 DOI 10.1111/j.1574-6941.2012.01441.x.

Driscoll N, Baia P, Fischer AT Jr, Brauer T, Klohnen A. 2008. Large colon resection and anastomosis in horses: 52 cases (1996–2006). *Equine Veterinary Journal* **40(4)**:342–347 DOI 10.2746/042516408X293529.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26(19)**:2460–2461 DOI 10.1093/bioinformatics/btq461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27(16)**:2194–2200 DOI 10.1093/bioinformatics/btr381.

Elli M, Colombo O, Tagliabue A. 2010. A common core microbiota between obese individuals and their lean relatives? Evaluation of the predisposition to obesity on the basis of the fecal microflora profile. *Medical Hypotheses* **75(4)**:350–352 DOI 10.1016/j.mehy.2010.03.022.

Ellis CM, Lynch TM, Slone DE, Hughes FE, Clark CK. 2008. Survival and complications after large colon resection and end-to-end anastomosis for strangulating large colon

volvulus in seventy-three horses. *Veterinary Surgery* **37(8)**:786–790 DOI 10.1111/j.1532-950X.2008.00449.x.

Fernandes KA, Kittelmann S, Rogers CW, Gee EK, Bolwell CF, Bermingham EN, Thomas DG. 2014. Faecal microbiota of forage-fed horses in New Zealand and the population dynamics of microbial communities following dietary change. *PLOS ONE* **9(11)**:e112846 DOI 10.1371/journal.pone.0112846.

Fischer S, Trefz P, Bergmann A, Steffens M, Ziller M, Miekisch W, Schubert JS, Köhler H, Reinhold P. 2015. Physiological variability in volatile organic compounds (VOCs) in exhaled breath and released from faeces due to nutrition and somatic growth in a standardized caprine animal model. *Journal of Breath Research* **9(2)**:027108 DOI 10.1088/1752-7155/9/2/027108.

French NP, Smith J, Edwards GB, Proudman CJ. 2002. Equine surgical colic: risk factors for postoperative complications. *Equine Veterinary Journal* **34(5)**:444–449 DOI 10.2746/042516402776117791.

Garner CE, Smith S, De Lacy Costello B, White P, Spencer R, Probert CS, Ratcliffe NM. 2007. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* **21(8)**:1675–1688 DOI 10.1096/fj.06-6927com.

Hackett ES, Embertson RM, Hopper SA, Woodie JB, Ruggles AJ. 2014. Duration of disease influences survival to discharge of Thoroughbred mares with surgically treated large colon volvulus. *Equine Veterinary Journal* **47(6)**:650–654 DOI 10.1111/evj.12358.

Harrison IW. 1988. Equine large intestinal volvulus a review of 124 cases. *Veterinary Surgery* **17(2)**:77–81 DOI 10.1111/j.1532-950X.1988.tb00281.x.

Hough R, Archer D, Probert C. 2018. A comparison of sample preparation methods for extracting volatile organic compounds (VOCs) from equine faeces using HS-SPME. *Metabolomics* **14(2)**:19 DOI 10.1007/s11306-017-1315-7.

Hudson JM, Cohen ND, Gibbs PG, Thompson JA. 2001. Feeding practices associated with colic in horses. *Journal of the American Veterinary Medical Association* **219(10)**:1419–1425 DOI 10.2460/javma.2001.219.1419.

Kaneene JB, Miller R, Ross WA, Gallagher K, Marteniuk J, Rook J. 1997. Risk factors for colic in the Michigan (USA) equine population. *Preventive Veterinary Medicine* **30(1)**:23–36 DOI 10.1016/S0167-5877(96)01102-6.

Leng J, Proudman C, Darby A, Blow F, Townsend N, Miller A, Swann J. 2018. Exploration of the fecal microbiota and biomarker discovery in equine grass sickness. *Journal of Proteome Research* **17(3)**:1120–1128 DOI 10.1021/acs.jproteome.7b00784.

Lozupone CA, Hamady M, Kelley ST, Knight R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology* **73(5)**:1576–1585 DOI 10.1128/AEM.01996-06.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2016. *cluster: cluster analysis basics and extensions.* R package version 2.0.4. *Available at https://cran.r-project.org/web/packages/cluster/index.html.*

McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE* **8(4)**:e61217 DOI 10.1371/journal.pone.0061217.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Michen PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2015. *vegan: community ecology package.* R package version 2.3-0. *Available at https://cran.r-project.org/web/packages/vegan/index.html.*

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2015. *nlme: linear and nonlinear mixed effects models.* R package version 3.1-122. *Available at https://cran.r-project.org/web/packages/nlme/index.html.*

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5(3)**:e9490 DOI 10.1371/journal.pone.0009490.

Proudman CJ, Hunter JO, Darby AC, Escalona EE, Batty C, Turner C. 2015. Characterisation of the faecal metabolome and microbiome of Thoroughbred racehorses. *Equine Veterinary Journal* **47(5)**:580–586 DOI 10.1111/evj.12324.

R Core Team. 2014. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. *Available at http://www.R-project.org/.*

Salem SE, Maddox TW, Berg A, Antczak P, Ketley JM, Williams NJ, Archer DC. 2018. Variation in faecal microbiota in a group of horses managed at pasture over a 12-month period. *Scientific Reports* **8(1)**:8510 DOI 10.1038/s41598-018-26930-3.

Schulz S, Dickschat JS. 2007. Bacterial volatiles: the smell of small organisms. *Natural Product Reports* **24(4)**:814–842 DOI 10.1039/b507392h.

Shannon CE. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**:379–423.

Snyder JR, Pascoe JR, Olander HJ, Spier SJ, Meagher DM, Bleifer DR. 1989. Strangulating volvulus of the ascending colon in horses. *Journal of the American Veterinary Medical Association* **195(6)**:757–764.

Suthers JM, Pinchbeck GL, Proudman CJ, Archer DC. 2013. Risk factors for large colon volvulus in the UK. *Equine Veterinary Journal* **45(5)**:558–563 DOI 10.1111/evj.12039.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. 2009. A core gut microbiome in obese and lean twins. *Nature* **457(7228)**:480–484 DOI 10.1038/nature07540.

Van Den Berg M, Hoskin SO, Rogers CW, Grinberg A. 2013. Fecal pH and microbial populations in Thoroughbred horses during transition from pasture to concentrate feeding. *Journal of Equine Veterinary Science* **33(4)**:215–222 DOI 10.1016/j.jevs.2012.06.004.

Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73(16)**:5261–5267 DOI 10.1128/AEM.00062-07.

Weese JS, Holcombe SJ, Embertson RM, Kurtz KA, Roessner HA, Jalali M, Wismer SE. 2015. Changes in the faecal microbiota of mares precede the development of post partum colic. *Equine Veterinary Journal* **47(6)**:641–649 DOI 10.1111/evj.12361.

Weiss SJ, Xu Z, Amir A, Peddada S, Bittinger K, Gonzalez A, Lozupone C, Zaneveld JR, Vazquez-Baeza Y, Birmingham A, Knight R. 2015. Effects of library size variance, sparsity, and compositionality on the analysis of microbiome data. *PeerJ PrePrints* **3**:e1157v1 DOI 10.7287/peerj.preprints.1157v1.

Wickham H. 2009. *ggplot2: elegant graphics for data analysis.* New York: Springer-Verlag.